# SETD1 HISTONE 3 LYSINE 4 METHYLTRANSFERASE COMPLEX COMPONENTS IN EPIGENETIC REGULATION

Patricia A. Pick-Franke

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Biochemistry and Molecular Biology
Indiana University

December 2010

Accepted by the Faculty of Indiana University, in partial
fulfillment of the requirements for the degree of Master of Science.

_____
David Skalnik, Ph.D., Chair

_____
Kristin Chun, Ph.D.

Master's Thesis
Committee

_____
Simon Rhodes, Ph.D.

## DEDICATION

This thesis is dedicated to my sons, Zachary and Zephaniah who give me great joy, hope and continuous inspiration. I can only hope that I successfully set a good example demonstrating that one can truly accomplish anything, if you never give up and reach for your dreams.

## ACKNOWLEDGEMENTS

I would like to thank my committee members Dr. Skalnik, Dr. Chun and Dr. Rhodes for allowing me to complete this dissertation. They have been incredibly generous with their flexibility. I must make a special thank you to Jeanette McClintock, who willingly gave her expertise in statistical analysis with the Cfp1 microarray data along with encouragement, support and guidance to complete this work. I would like to thank Courtney Tate for her ceaseless willingness to share ideas, and her methods and materials, and Erika Dolbrota for her generous instruction as well as the name of a good doctor. I would also like to acknowledge the superb mentorship of Dr. Jeon Heong Lee, PhD and the contagious passion and excitement for the life of science of Dr. Clark Wells.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## ABBREVIATIONS

| | |
|---|---|
| ANOVA | analysis of variation |
| BME | β-mercaptoethanol |
| CFP1 | CXXC Finger Protein 1 |
| CGI | CpG dinucleotide island |
| CPF | cleavage and polyadenylation factor |
| CTD | carboxy-terminal heptad repeat domain |
| DAPI | 4'-6-Diamidino-2-phenylindole |
| DAVID | database for annotation, visualization and integrated discovery |
| DMEM | Dulbecco's modified Eagle medium |
| DNMT | DNA methyltransferase |
| DPC | Days post coitum |
| DTT | dithiothreitol |
| EDTA | ethylene diamine tetraacetic acid |
| EGTA | ethylene glycol tetraacetic acid |
| ES | embryonic stem |
| FACS | flourescent activated cell sorting |
| FBS | fetal bovine serum |
| FCS | fetal calf serum |
| GO_BP | Gene Ontology Biological Process |
| H3K4 | histone 3 lysine 4 |
| HCF-1 | Host cell factor 1 |
| HEK | human embryonic kidney |
| HEPES | 4-(2-hyroxyethyl)-1-piperazineethanesulfonic acid |
| HET | heterozygote |
| LIF | leukemia inhibitory factor |
| MAS5 | microarray algorithm suite version 5 |
| PANTHER | protein analysis through evolutionary relationships |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| Pen/Strep | penicillin/streptomycin |
| PI | propidium iodide |
| PMSF | phenylmethysulfonyl flouride |
| Pol II | RNA polymerase II |
| PP1 | protein phosphatase I |
| RRM | RNA recognition motif |
| RT | room temperature |
| SDS | sodium dodecyl sulfate |
| TBST | Tris buffered saline with Tween |
| Tris | tris (hydroxymethyl) aminomethane |
| Wdr82 Het | B-galactosidase disrupted Wdr82 gene containing ES cell line |
| WT | wild type |

# CHAPTER 1: INTRODUCTION

## 1.1 Epigenetic Mechanisms

Epigenetics is defined as heritable patterns of gene expression that occur without changes to DNA sequence. Epigenetic regulation is crucial for the process of cellular differentiation and development as evidenced by distinctly different sets of gene transcripts found in diverse tissues despite carrying identical genetic information (Bartolomei & Tilghman, 1997; Felsenfeld & Groudine, 2003). Epigenetic control is largely facilitated by regulation of transcription through dynamic remodeling of chromatin structure.

The basic unit of chromatin is the histone core, an octomer of four histone proteins, H2A, H2B, H3, and H4 and approximately 147 base pairs of DNA wrapped around it. Actively transcribed chromatin is found in a more open configuration called euchromatin in contrast to non-transcribed heterochromatin found in a tightly closed configuration. Specific post-translational modifications including acetylation, phosphorylation, sumoylation, ubiquitination and methylation of specific amino acids on histone proteins have been associated with chromatin status (Felsenfeld & Groudine, 2003; Grewal & Moazed, 2003; Jaenisch & Bird, 2003). For example, H3 methylated at lysine 4 (H3K4) is associated with euchromatin and H3 methylated at lysine 9 (H3K9) is associated with heterochromatin. Many of the enzyme complexes associated with these specific post-translational modifications have been identified, however the specific events that lead to the targeting of these enzymes are still largely unknown.

Methylation of mammalian DNA provides an additional mechanism for epigenetic control. Cytosine methylation, covalent attachment of a methyl group at the C5 position of cytosine, occurs usually in the context of CpG dinucleotides and is largely associated with gene repression (Bird, 2002). Methyl cytosine spontaneously deaminates to thymine, whereas unmethylated cytosine converts to uracil. If left unrepaired, thymine would be recognized as a valid base and DNA replication would result in a shift to adenosine instead of guanine at that location. It is believed that as a consequence of DNA repair efficiency (lower for thymine vs. uracil) following spontaneous cytosine deamination, CpG dinucleotide distribution is found at only ~10% of the expected frequency throughout the mammalian genome (Bird, 2002; Antequera & Bird, 1993; Larsen et al., 1992). However, CpG dinucleotides are found at the expected frequency in clusters of non-methylated CpG dinucleotides denoted CpG islands (CGI). Unmethylated CpG dinucleotides are enriched in regions near promoters of 60-70% of all human genes (Lander, 2001). CGIs colocalize with the promoters of all human genes and approximately 40% of those found expressed in a tissue restrictive manner (Larsen et al., 1992; Zhu et al., 2008). Several transcripts initiate from intragenic CGIs and are expressed during specific developmental stages (Macleod et al., 1998; Kleinjan et al., 2004).

The *de novo* DNA methyltransferases (Dnmt) Dnmt3a and Dnmt3b establish cytosine methylation patterns during early development (Okano et al., 1999) and Dnmt1, which preferentially acts on the hemi-methylated DNA replication product, is largely responsible for maintaining the pattern (Yoder et al., 1997). Lower eukaryotes do not utilize cytosine methylation as a mechanism for epigenetic regulation, however this

machinery is critical for mammalian development, as ablation of any of the Dnmt genes in mice is lethal (Lei et al., 1996; Okano et al., 1999). Alteration of cytosine methylation patterns is frequently observed in cancers (Baur et al., 1999; Guo et al., 2000; Issa et al., 1997; Hennessy et al., 2003) and leads to disease states such as immunodeficiency (Xu et al., 1999). Dynamic cytosine methylation plays an important role in gametogenesis, X chromosome inactivation and genomic imprinting and changes in DNA methylation along with histone modifications are necessary for in vitro differentiation of embryonic stem (ES) cells (Bestor, 2000; Okano & Li, 2002; Yokochi & Robertson, 2002). ES cells carry a relatively open chromatin structure and the orderly restriction of lineage potential during development is associated with progressive accumulation of heterochromatin and restriction of gene expression (Akashi et al., 2003; Attema et al., 2007).

### 1.2 H3K4 Methylation and Enzymatic Complexes

Lysine methylation occurs in either a mono-, di- or trimethylated form (Santos-Rosa et al., 2002). Trimethylation of H3K4 is largely associated with actively or recently transcribed genes (Ng et al., 2003a) although H3K4 trimethylation has been identified in non-transcribed developmental genes in embryonic stem cells (Bernstein et al., 2006). The H3K4 trimethylation mark is primarily found at promoter and transcription start sites within the gene (Guenther et al., 2007). Dimethylation is predominantly identified in the middle, and mono-methylation is predominantly identified toward the 3' end of actively transcribed genes (Pokholok et al., 2005), although the function of mono- and di-methylation remains unknown. Interestingly, under transcriptional stress induced by

depletion of nucleoside triphosphates, trimethylation appears at the 3' end of the coding region, suggesting specific signaling factors are involved in regulating H3K4 methyltransferase activity (Ng et al., 2003a).

In yeast, a single multi-component enzymatic complex known as Set1/COMPASS performs H3K4 methylation. Set1/COMPASS is comprised of a large scaffold protein, called Set1, containing a conserved catalytic C-terminal SET domain (named from the *Drosophila* genes Su(var) 3-9, Enhancer of zeste, and trithorax) (Jenuwein, 1998) as well as an N-terminal RNA recognition motif (RRM). Seven additional proteins belong to the complex, most of which are required for efficient H3K4 methylation. Interestingly, deletion of Set1 abolishes all H3K4 methylation in *Saccharomyces cerevisiae,* yet it is not lethal (Boa et al., 2003; Briggs et al., 2001; Roguev et al., 2001). However, Set1 deletion affects numerous important functions such as growth, transcriptional activation, repression and elongation, ribosomal DNA silencing, meiotic differentiation, DNA repair, chromosome segregation, and cell wall organization as well as telomere length regulation. For example, deletion of Set1 results in aberrant position dependent gene silencing at telomeres, the mating type locus and ribosomal DNA in *Saccharomyces cerevisiae* (Briggs et al., 2001; Corda et al., 1999; Krogan et al., 2002; Nislow et al., 1997). Defects in these cellular processes following deletion of Set1 suggests Set1/COMPASS is targeted to genes within these functional pathways, although the mechanism for targeting remains elusive.

Set1 associates with the coding regions of highly transcribed RNA polymerase II (Pol II) genes, and the presence of Set1 generally correlates with H3K4 trimethylation (Ng et al., 2003c). Set1 associates with Pol II when it's carboxyl-terminal heptad repeat

domain (CTD) is phosphorylated at Serine 5, a mark of transcription initiation (Ng et al., 2003c), and this association is dependent on the Paf1 complex and the mono-ubiquitination of lysine 123 on histone H2B by Rad6/Bre1 (Krogan et al., 2003; Mueller & Jaehning, 2002). The Paf1 complex is required for Rad6/Bre1 activity, but not targeting of the ubiquitin protein ligase to promoters (Krogan et al., 2003). H2B ubiquitination is a requirement for di- and trimethylation by Set1, but not mono-methylation (Hwang et al., 2003; Ng et al., 2003b; Dehé et al., 2005; Shahbazian et al., 2005).

Only one subunit of the Set1/COMPASS complex, Swd2, is essential for survival in yeast (Cheng et al., 2004; Dichtl et al., 2004; Roguev et al., 2004). Interestingly, Swd2 is reported to be a component of the cleavage and polyadenylation factor (CPF) necessary for mRNA processing and transcription termination in addition to the Set1/COMPASS complex (Cheng et al., 2004; Dichtl et al., 2004; Roguev et al., 2004). Recruitment to CPF is dependent on the Ref2 subunit, an RNA-binding protein known to stimulate processing at weak poly(A) sites (Cheng et al., 2004). Swd2 mutations generated with error prone PCR mutagenesis resulted in variable effects depending on which gene was evaluated in *Saccharomyces cerevisiae,* suggesting the function of Swd2 in linking H3K4 methylation and Pol II termination may be gene dependent (Cheng et al., 2004).

Other components of the yeast Set1/COMPASS complex include Sgh1 (Cps 15), Swd1 (Cps 50), Swd3 (Cps 30), Bre2 (Cps60), Sdc1 (Cps25) and Spp1 (Cps 40) (Roguev et al., 2001; Nagy et al., 2002; Miller et al., 2001). Swd1 and Swd3, which remain associated with each other in the absence of Set1, are essential for the integrity of the

COMPASS complex and stability of Set1 (Morillon et al., 2005; Roguev et al., 2001; Schneider et al., 2005). Bre2, Sdc1 and Spp1 are required for proper trimethylation levels and location within the transcribed genes (Dehé et al., 2006). Mono- and di-methylation levels are influenced by Bre2/Sdc1, but not Spp1 (Dehé et al., 2006). Spp1 also plays a role in Set1 stability as cells lacking Spp1 show a clear reduction in the level of Set1 (Nagy et al., 2002; Roguev et al., 2001; Schneider et al., 2005). These proteins are also highly conserved in organisms from yeast to humans (Roguev et al., 2001; Roguev et al., 2003).

In contrast to yeast, humans have multiple H3K4 methyltransferase complexes including Setd1A, Setd1B, Mll1-4, Smyd 3 and Set7/9. The family of H3K4 methyltransferases is widely expressed and appears to have non-redundant functions, as deletion or disruption of a single member can result in either death or disease. For example, Mll1 plays a role in development and hematopoiesis as the translocation of the Mll1 gene is observed in multiple types of leukemia (Tenney & Shilatifard, 2005). Genetic disruption of Mll1 or Mll2 results in embryonic lethality (Glaser et al., 2006; Ayton et al., 2001). Although Mll1 catalyzes H3K4 methylation of two specific genes, HoxA9 and HoxC8 (Ansari et al., 2008), and Mll2 has been shown to be required for male and female gamete development in mice (Andreu-Vieyra et al., 2010), the specific mechanism and targeting of each H3K4 methyltransferase remains to be determined.

### 1.3 Setd1A and Setd1B

Setd1A and Setd1B are the most closely related to the yeast Set1, including association with all six non catalytic components comprised of Cfp1 (Spp1 analog),

Rbbp5 (Swd1 analog), Ash2 (Bre2 Analog), Wdr5 (Swd3 analog) and Wdr82 (Swd2 analog). In contrast, Mll1 and Mll2 associate with Cfp1 (Ansari et al., 2008) but not Wdr82 (Lee et al., 2010), and Mll3 and Mll4 associate with Wdr82 but not Cfp1 (Lee et al., 2010). The two Setd1 methyltransferases exhibit both overlapping and non-redundant properties. Human Setd1A and Setd1B are 85% identical throughout the catalytic SET and post-SET domains and 61% identical throughout the N-terminal region that includes a RRM domain (Lee et al., 2007). The central region contains only 46% identity (Lee et al., 2007). Inducible expression of the C-terminus of either Set1dA or Set1dB decreases steady-state levels of both endogenous Set1dA and Set1dB proteins, but does not alter the expression of the associated components of the Set1 complexes (Lee et al., 2007). However, nuclear localization of these two methyltransferases are largely non-overlapping and mostly confined to euchromatin speckles as seen in confocal microscopy, suggesting distinct gene targets (Lee et al., 2007). In addition, a host cell factor 1 (HCF-1) binding motif unique to the Setd1A central region interacts with HCF-1 *in vivo* (Lee et al., 2007; Tyagi et al., 2007), but is not found in Setd1B. Interestingly, HCF-1 mediated recruitment of Mll1 to E2F responsive promoters by E2F1 promote cell cycle progression (Takeda et al., 2006; Tyagi et al., 2007). The significance of HCF-1 in the cell cycle and its association with Setd1A but not Setd1B is consistent with separate and distinct roles for the two Setd1 proteins.

As in yeast, human Setd1 H3K4 methyltransferases localize with Pol II (Ng et al., 2003a; Lee & Skalnik, 2008), and are highly associated with the 5' end of actively transcribed genes. However, in contrast to the yeast requirement for association with the

Paf1 complex for Set1 interaction with Pol II, the human Setd1A and Setd1B complexes associate with Pol II by interaction with Wdr82 (Lee & Skalnik, 2008).

### 1.4 Wdr82

Set1/COMPASS subunits of budding and fission yeast have been extensively characterized, however much less is known about the mammalian homologue. Only 4 publications provide specific information on Wdr82. Wdr82 is one of three COMPASS subunits made up of multiple WD-40 motifs, a motif with circular propeller-like structures which can interact sequentially or simultaneously with several different proteins. Wdr82 associates with the RRM domain of Setd1A and directly recognizes phosphorylated serine 5 in the CTD of the Pol II large subunit, but does not associate under conditions of unphosphorylated serine 5, or phosphorylation at serine 2 found during transcription elongation (Lee & Skalnik, 2008). Depletion of Wdr82 by small interfering RNA leads to decreased Setd1A expression and occupancy at transcription start sites coupled with reduced H3K4 trimethylation (Lee & Skalnik, 2008). Recently, Wdr82 has been identified in association with other protein complexes including a protein phosphatase 1 (PP1) containing complex, a chaperone containing Tcp1 complex and other uncharacterized proteins (Lee et al., 2010) indicating a role in multiple functions.

### 1.5 Cfp1

Similar to Wdr82, a decrease in Cfp1 protein levels leads to reduced Setd1A protein levels, however, global levels of H3K4 trimethylation increases (Tate et al.,

2010). In yeast, the homolog Spp1 is not required for survival, however deletion of Cfp1 results in embryonic lethality in mice as homozygously disrupted Cfp1 allele mice could not be recovered from heterozygous mice matings. (Carlone & Skalnik, 2001). Blastocysts lacking Cfp1 appear normal and appear to implant but fail to gastrulate. ES cell lines lacking Cfp1 are viable, but present an altered phenotype with an increased rate of apoptosis and an inability to differentiate (Carlone et al., 2005; Tate et al., 2009b). ES cells lacking Cfp1 also exhibit a dramatic decline in global cytosine methylation and altered histone modifications associated with heterochromatin (Carlone et al., 2005; Lee & Skalnik, 2005; Tate et al., 2010) as well as a global reduction in protein synthesis despite evidence of increased transcripts for some genes (Butler et al., 2008; Butler et al., 2009).

Cfp1 exhibits a unique binding specificity to DNA sequences containing unmethylated CpG motifs (Voo et al., 2000) and a search for proteins common to all CGI's showed a high level of enrichment of Cfp1 (Thomson et al., 2010). A reduction of H3K4 trimethylation in CGIs was reported in Cfp1 depleted NIH 3T3 cells suggesting genes associated with CGIs are likely targets for Cfp1 interacting methyltransferases. In addition to binding unmethylated CpG motifs, Cfp1 interacts directly with Dnmt1, the maintenance DNA methyltransferase, independently of association with the Setd1 H3K4 methyltransferase complexes (Butler et al., 2008), suggesting Cfp1 may enable crosstalk between the histone and DNA methylation epigenetic marks. Remarkably, transfection studies in ES cells reveal that either the carboxyl half which has a SET binding motif, or the amino terminal half containing the DNA binding motif of Cfp1, is sufficient to rescue defects of Cfp1 depleted ES cells, provided the binding motifs within are not altered

(Tate et al., 2010). Three fragments of the Cfp1 protein, the carboxyl half and two within the amino half, are sufficient to interact with Dnmt1 (Butler et al., 2008). Taken together, these data suggest Cfp1 plays a critical role in regulating chromatin configuration and consequent transcriptional expression through epigenetic mechanisms and may play a role in developmental changes in DNA methylation.

The following studies further investigate the impact of altering protein levels and protein configuration for Setd1 complexes and the critical components, Wdr82 and Cfp1.

# CHAPTER 2: MATERIALS AND METHODS

## 2.1 Cell culture conditions and staining procedures

### *2.1.1 ES cell culture*

Undifferentiated ES cells were grown on tissue culture plates coated with 0.1% gelatin in media containing high glucose Dulbecco's modified Eagle Medium (DMEM) (Gibco BRL), 20% fetal bovine serum (FBS)(Gibco), 100 units Pennicillin/Streptomycin (Pen/Strep) (Invitrogen), 2 mM L-Glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 0.2% conditioned medium from leukemia inhibitory factor (LIF) generating chinese hamster ovary cells and 100 mM BME solution made with 100 nM β-mercaptoethanol (BME), 0.025% 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 1% Hank's balanced salt solution (Invitrogen).

### *2.1.2 Differentiation of ES cells*

To induce differentiation, asynchronous exponentially growing cells were trypsinized and transferred to conical tubes for centrifugation and removal of media containing LIF. Cells were washed two times and re-suspended in growth media without LIF, and plated at $1 \times 10^7$ cells per plate without gelatin. ES cells were monitored daily for morphological changes on an inverted light microscope (10x) and media was changed on day 3 being careful not to disrupt embroid-like bodies if present.

*2.1.3 β-galactosidase staining of ES cells*

Wild type (WT) ES cells and an ES cell line which has a β-galactosidase (LacZ) gene inserted between exons 3 and 4 of the Wdr82 gene, obtained from the Mutant Mouse Regional Resource Center (AZ0280) (designated hereafter as Wdr82 Het), were grown in ES cell media as described above. Media was removed from asynchronous exponentially growing cells, followed by a wash with PBS and cells were fixed with X-gal fix buffer (0.1 M phosphate buffer pH 7.3 supplemented with 5 mM ethylene glycol tetraacetic acid [EGTA] pH 7.3, 2 mM $MgCl_2$ and 0.2% gluteraldehyde) for 15 minutes at room temperature (RT). The cells were washed twice with X-gal wash buffer (0.1 M phosphate buffer pH 7.3 supplemented with 2 mM $MgCl_2$) followed by addition of X-gal staining buffer [0.1 M phosphate buffer pH 7.3 supplemented with 2 mM $MgCl_2$, 5 mM potassium ferrocyanide ($K_4Fe(CN)_6$-$3H_2O$) and 5 mM potassium ferricyanide ($K_5Fe(CN)_6$) with 1 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal)] with just enough liquid to cover the cells and placed inside a humidified darkened chamber at 37˚C overnight. The following day, X-gal staining buffer was removed, cells were washed with X-gal wash buffer and observed for blue color indicating β-galactosidase activity with the inverted light microscope under 10x magnification.

*2.1.4 Human embryonic kidney (HEK) 293 T-rex growth media and expression induction*

Stable doxycycline inducible HEK 293T-rex cell lines transfected with various Setd1B cDNA fragments, full length Setd1A, full length Setd1B or vector were cultured in DMEM tissue culture media containing 10% fetal calf serum (FCS) (Gibco BRL), 50 units Pen/Strep (Invitrogen), 2 mM L-Glutamine (Invitrogen), 20 µl of 25 mg/ml

blasticidin, and 500 µl hygromycin. To induce expression of cDNAs, 1 mg/ml doxycycline was added to cell media. The cell lines were supplied by Dr. Jeon-Heong Lee.

*2.1.5 HEK 293 T-rex cell growth counting*

Stably transfected HEK 293 T-rex cell lines were plated in duplicate at 5 x $10^4$ cells per well and induced to express the vector DNA by of doxycycline to the growth media. At 24 hour intervals for four days, cells were collected and counted in triplicate by hemocytometer under the light microscope. To maximize complete collection of all cells for counting, all fluid and/or cells were transferred into a single conical tube for each step in the cell collection process, followed by centrifugation to concentrate and re-suspension for counting. Briefly, growth media was removed, cells washed once with PBS followed by addition of trypsin and two additional well washes with PBS. Cells were spun at 4˚C to form a pellet and re-suspended in 1 ml of media lacking doxycycline. Trypan Blue was added to cells for visualization and counting by hemocytometer. Only viable cells were counted.

*2.1.6 Cell cycle analysis by propidium iodide*

Asynchronous exponentially growing induced HEK 293 T-rex cells cultured in media containing doxycline were trypsinized, counted and adjusted to 5 x $10^5$ cells/tube followed by two washes in cold PBS. Cells were re-suspended and fixed in cold 70% ethanol and stored at -20˚C overnight or until analysis. Cells were washed twice in cold PBS and re-suspended in fresh propidium iodide (PI) buffer (0.5 mg/ml propidium

iodide, 0.1% Triton-X-100, 1 mM ethylenediaminetetraacetic acid [EDTA], 4 mg/ml RNAase A in PBS pH 7.4, 100 µg/ml) for DNA staining. Cells were maintained on ice protected from light until analysis on the fluorescent antibody cell sorter (FACS) using standard PI programming for cell cycle analysis on FACScan flow cytometer (Becton Dickinson). Multiple biological samples were evaluated, and duplicate technical samples were stained and analyzed within each biological sample.

## 2.2 Protein analysis by western blot

### *2.2.1 Cellular extracts*

WT and Wdr82 Het ES cells were harvested for protein analysis. For each sample, ES cells from two to three 15 cm culture plates were removed and placed into a 15 ml conical tube, followed by centrifugation at 1200 rpm for 2 minutes at 4˚C. Supernatant was removed and cells were re-suspended in 4˚C hypotonic buffer (10 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl supplemented with 1 µg/ml each leupeptin, aprotinin, pepstatin, 1 mM phenylmethysulfonyl fluoride (PMSF) and 1 mM Dithiothreitol (DTT) for protease inhibition and placed on ice for 10 minutes, followed by homogenation with 10 strokes of Dounce homogenizer and centrifugation at 16,000 x g for 10 minutes at 4˚C. Supernatant containing cytosolic components was transferred to a separate tube and cell pellets were re-suspended in cyto-skeletal buffer (10 mM 1,4-piperazinediethanesulfonic acid [PIPES] pH 7.0 300 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 1 mM EGTA, and 0.5% Triton X-100) supplemented with 1µg/ml each leupeptin, aprotinin, pepstatin and 1 mM PMSF and 1 mM DTT and incubated on ice for 10 min, followed by homogenization with 10 strokes in Dounce glass homogenizer. The

solubilized nuclear fraction was separated by centrifugation at 16,000 x g for 10 minutes at 4˚C.

*2.2.2 SDS Page electrophoresis and western blot*

Nuclear extracts were mixed with 4x concentrated Laemmli sample buffer (20% glycerol, 2% sodium dodecyl sulphate [SDS], 5% BME, 0.1% bromophenol blue and 62.5 mM tris[hydroxymethyl]aminomethane [Tris], pH 6.8) and heated for 5 minutes at 95˚C then placed on ice to cool. Protein samples (10-15 µl) were subjected to electrophoresis on 8% gel and transferred to nitrocellulose membranes (Amersham, GE Healthcare). The membranes were blocked with 5% milk in Tris buffered saline with Tween (TBST) for one hour and washed three times with TBST before being probed with primary antiserum rabbit polyclonal anti WDR82 (WD518A), 1:1000 or mouse monoclonal β-actin 1:1000 (Sigma-Aldrich) for 1 hour at RT or overnight at 4˚C. Membranes were washed three times with TBST and incubated with horseradish peroxidase labeled secondary antibody 1:1000 for 1 hour at RT. Membranes were washed three times with TBST and signal was detected with ECL kit (Amersham, GE Healthcare).

**2.3. Embryo collection and analysis**

*2.3.1 Timed pregnancy*

C57BL/6J mice heterozygous for disrupted Wdr82 gene ($Wdr82^{(+/-)}$) mice were mated. Two $Wdr82^{(+/-)}$ females were caged overnight with one $Wdr82^{(+/-)}$ male. The females were checked for vaginal plugs the following morning and the appearance of

vaginal plug was designated as day 0.5 of gestation. Pregnant females were sacrificed at day 9.5 to 11.5 by $CO_2$ inhalation and embryos were harvested, washed in PBS pH 7.3 and observed for phenotypic defects. The care and use of animals in these studies were approved by the Indiana University School of Medicine Institutional Animal Care and Use Committee.

*2.3.2 Embryo Digestion*

Embryos obtained from $Wdr82^{(+/-)}$ matings were digested in cell lysis buffer (10 mM Tris, 100 mM EDTA, 0.5% SDS pH 8.0) plus 20 mg/ml proteinase K (Roche) and incubated at 55˚C for 4-24 hours depending on the size of embryo. Solution was put on ice to cool and incubated with 5 µl RNAse A (10 mg/ml) for 15 minutes at 37˚C and returned to ice to cool. To precipitate protein, 7.5 M ammonium acetate (200 µl) was added, samples were vortexed for 20-30 seconds and put on ice for 5 to 30 minutes. Tubes were spun in a micro-centrifuge at maximum speed at 4˚C for 15 minutes and supernatant was transferred to a fresh tube for DNA precipitation.

*2.3.3 DNA precipitation*

To precipitate DNA, 70% volume/volume isopropanol was added to microfuge tube containing DNA supernatant and inverted to mix several times followed by a spin in micro-centrifuge at maximum speed for 1 minute at RT to form a DNA pellet. Supernatant was poured off and the pellet was washed with 70% ethanol, followed by a wash with ice cold 100% ethanol and allowed to air dry. DNA pellets were re-suspended in 50 µl sterile distilled water.

*2.3.4 Genotyping by Polymerase Chain Reaction (PCR)*

Multiplex PCR was performed with 1 µl re-suspended DNA in a 50 µl reaction mixture containing 0.2 mM of each deoxynucleotide triphosphate (dNTP), 50 pmol each of forward primer P8a and backward primers P7 and P2a, 1 unit of Taq DNA polymerase (Roche) in 10x buffer supplied by the manufacturer (10x PCR buffer with $MgCl_2$). Cycling conditions for PCR included heat denaturation at 94˚C for 4 minutes, followed by 25 cycles of 94˚C for 1 minute, 62˚C for 1 minute, and 72˚C for 1 minute, and a final elongation step at 72˚ C for 10 minutes. Previously characterized mouse tail and ES cell DNA samples were used as controls. PCR products (25 µl) were run on a 1.5% agarose gel with ethidium bromide. The size was determined by a 100 bp DNA standard ladder.

Primer sequences used:

P2a 5'- CAGGTTTGCCCTGTAGATGCATGAG - 3'

P8a 5'- ACTGAGCCAGCCTTTCATGGCAGC - 3'

P7 5'- TACCCCAGACCTTGGGACCACC - 3'

**2.4 ES cell line microarray analysis**

*2.4.1 Sample preparation and hybridation*

Murine ES cell lines, including one wild type (WT), two CFP1 disrupted (N1 and N2), and one N1 ES cell line transfected with human CFP1 cDNA (Rescue) were grown in media containing LIF. Asynchronous exponentially growing cells were processed for total RNA by Courtney Tate. Four biological samples from each cell line were handled in parallel to reduce non-random error and supplied to the Indiana University Core Center for Genomics facility for analysis. Briefly, total RNA samples were labeled using the

Affymetrix standard One-Cycle protocol and hybridized to Affymetrix Mouse Genome 430 2.0 GeneChips® according to standard Affymetrix protocols. GeneChips were scanned with an Affymetrix Model 3000 scanner controlled by GCOS software (Affymetrix). Microarray Suite version 5 (MAS5) signals and detection call were generated and scaled using a target of 1000.

*2.4.2 Data analysis*

The Affymetrix MAS5 algorithm was used to analyze the hybridization intensity data from each array. The MAS5 data were imported into Partek Genomics Suite (Partek, Inc., St. Louis, MO) software and was log transformed (base 2) followed by evaluation of overall characteristics for quality assessment with principal components analysis (PCA) and hierarchical clustering. Gene annotations were supplied by Affymetrix (www.Affymetrix.com) from the most recent update (August 10, 2010).

Probe sets were then removed unless a determination of "present" was made in at least 50% of the replicates within a minimum of one of the experimental cell line groups (McClintick & Edenberg, 2006).

To view initial global patterns of gene expression changes and rescue, a one-way analysis of variation (ANOVA) was performed using all four cell lines. The top 100 genes were selected for hierarchical clustering.

A two-way ANOVA was performed by genotype (average of WT and Rescue designated as WT genotype and average of N1 and N2 designated as NULL phenotype) and samples (ES cell line within genotype) plus relevant contrasts. Euclidian distance, a measure employed to identify differences, was used to cluster genotype samples, and

Pearson's Dissimilarity Coefficient, a measure employed to identify similarities, was used to cluster the gene probes.

Contrasts were made for N1 vs. WT, N2 vs. WT, Nl vs. Rescue, Rescue vs. WT, and N1 vs. N2. Statistical determination for p-values, false discovery rate (FDR) designated as q-value and fold changes for each of these contrasts were calculated. FDRs were calculated using the method from Storey and Tibshirani (Storey & Tibshirani, 2003). The fold change calculation was based on the untransformed signal. Statistical analysis was performed by The Center for Medical Genomics at Indiana University-Purdue University Indianapolis (IUPUI).

Probe lists for functional analysis were generated using two-way ANOVA contrasts and imported into the database for annotation, visualization and integrated discovery (DAVID) bioinformatics resources (http://david.abcc.ncifcrf.gov/). First, data was filtered to remove probes not called present in at least 50% of any cell line tested. Due to the high number of genes identified as statistically different, a filter was applied to reduce the number of genes submitted to DAVID and increase the confidence in functional annotation analysis. Probes statistically significant in N1 vs. WT, N2 vs. WT and N1 vs. Rescue contrasts at $p<.0001$ and consistent fold change +/- 1.5 were identified and separated into up-regulated and down-regulated gene lists and subsequently submitted to DAVID for analysis (October, 2010). Multiple probes representing the same gene were removed to eliminate over-representation.

To evaluate genes that did not rescue, a filter of $p<.05$ was initially evaluated in all N1 vs. WT, N2 vs. WT and N1 vs. Rescue contrasts where N1 and N2 vs. WT were up-regulated but N1 vs. Rescue was down-regulated and vice versa. After submission to

DAVID for analysis, the lists were further filtered with fold change cutoff +/- 1.5 and resubmitted. Gene Ontology Biological Process was evaluated by DAVID and Affymetrix annotation.

# CHAPTER 3: RESULTS

## 3.1 Wdr82 is required for early embryonic development

Previously, cross-breeding of heterozygous Wdr82 mice in our lab resulted in no viable mice that were homozygous for the disrupted Wdr82 allele, indicating that Wdr82 is essential for murine development. To determine the time in development that Wdr82 is required, we collected embryos from heterozygous Wdr82 mice matings at day 9.5 to 11.5 days post coitum (dpc), the earliest time point that embryo and placental material can be clearly separated. The embryos were observed for phenotypic defects and none were observed (data not shown). To identify the genotype of embryos, multiplex PCR was performed using one common forward primer and two backward primers, with one being unique to the wild type Wdr82 gene located downstream from the disrupted gene location resulting in a higher base pair PCR product, and the second located within the β - galactosidase disruption cassette, resulting in a smaller product (Figure 1).

After analysis of 74 embryos, no homozygous disrupted Wdr82 genotypes were identified (Table 1). Interestingly, an overabundance of heterozygous genotype was observed and this finding is consistent with the ratio of WT and Wdr82 hets observed in full term pups (data not shown). To look earlier in development, uteri from Wdr82 het cross-bred females were collected at day 6.0 to 7.5 dpc. At this time point, embryos cannot be easily separated from the uterine deciduum, therefore uteri were sectioned and stained for histological analysis. Over thirty-five uteri from multiple matings were analyzed and no phenotypical differences were observed (data not shown). These results

suggest Wdr82 is required in development prior to day 9.5 of gestation and lack of Wdr82 may result in a pre-implantation defect.

### 3.2 The LacZ gene trap is functional

LacZ gene trap insertions in ES cells should express detectable levels of the β-galactosidase reporter in cells with maintenance and/or differentiation culture conditions. However, the fusion transcript may not encode a functional β-galactosidase fusion product. For example, insertions that capture the N-terminal signal sequence of a secreted or membrane protein are inactive as reported by Skarnes et al. (Skarnes et al., 1995). This is likely due to the inability of the enzyme to form a tetramer which is required for activity. Wdr82 is found in the cytoplasm as well as the nucleus and must pass through the nuclear membrane. In order to determine if the β-galactosidase-Wdr82 fusion product was functional, the Wdr82 het ES cell line used to generate the heterozygous Wdr82 mouse was examined. WT and Wdr82 het ES cells were stained with X-gal and observed for blue color indicating β-galactosidase activity. The characteristic blue color was observed in Wdr82 het cells, and not WT, however not every cell in the Wdr82 het dish turned blue. X-gal staining was clearly absent in wild type ES cells, indicating the inserted β-galactosidase gene trap is functional (Figure 2).

### 3.3 Wdr82 het ES cells have reduced Wdr82 protein levels and exhibit early induction of differentiation

To further characterize the Wdr82 het ES cell line, Wdr82 protein levels were analyzed by western blot. Nuclear lysates demonstrated a reduced amount of Wdr82

protein compared to WT ES cells (Figure 3). Consistently, when utilizing visual appearance as a standard for cell quantity, the Wdr82 het cell line exhibited higher levels of actin by western blot, a finding that may indicate a slightly higher rate of cell division which should be evaluated further (data not shown).

ES cell differentiation was assessed by washing and re-plating both WT and Wdr82 het ES cells in media without LIF to induce differentiation. Typically, when induced to undergo differentiation, murine ES cells form embroid-like bodies appearing as free floating spherical clusters. These spherical clusters eventually attach to the non-gelatin coated culture plate and begin to generate outgrowth by approximately day 3. The Wdr82 het ES cells formed embroid-like bodies similar to WT, however attachment and outgrowth occurred early in day 2 (Figure 4). Taken together with altered Wdr82 levels, this suggests a difference in differentiation dynamics may result from altered Wdr82 levels.

### 3.4 Setd1A and Setd1B play a role in cell division

Since Setd1A and Setd1B display largely non-overlapping nuclear localization, we speculated that they interact with non-redundant functional gene targets. It was speculated that since HCF-1 interaction with Setd1A is involved in cell cycle progression and unique to Setd1A, an evaluation of growth and cell cycle differences could elucidate the structural region responsible for functional differences between the Setd1 proteins. Exogenous expression of progressively larger N-terminal truncations of Setd1B cDNA should help identify the structural component unique to Setd1B. N-terminal truncations in this study are named by the nucleic acid in which the N-terminal truncation fragment

initiates, including B-676 which does not contain the RRM motif, and B-2482, B4366, B-4858, B-4981 or B-5266. All truncations contain the carboxyl terminal SET and post-SET domains. All cell lines expressing either truncated or full length Setd1 proteins demonstrated significant differences in cell numbers compared to vector at day 4 ($p<0.0001$) indicating both Setd1A and Setd1B play a role in cell division (Figure 5). Compared to Setd1A, full length Setd1B and B-676 were significantly different ($p<0.005$) but B-2482 through B-4981 were not, suggesting the region of functional difference may be in the area coded between nucleotides 676 through 2482 (Figure 6). Interestingly, B-5266, containing only SET and post-SET domains was significantly different than both Setd1A and Setd1B in addition to vector, suggesting the C-terminal SET domain is sufficient to interact with binding partners influencing cellular division.

The same clones were analyzed for cell cycle changes by PI analysis. PI binds DNA and the level of PI signal identified by flow cytometry correlates with DNA status. The signal is normalized to identify a single set of chromosomes (G1 phase), completion of replication with two sets of chromosomes (G2 phase) and intermediate signals identifying cells in the process of replication (S phase). PI cell cycle analysis revealed a significantly higher percentage of cells in G1 phase and fewer cells in S phase for cell lines expressing Set1dA, Setd1B, B-2482, B-4855 and B-4981 compared to vector ($p<0.01$), but no statistical difference was identified between Setd1A and Setd1B (Figure 7). This data suggests Setd1A and Setd1B both play a significant role in the cell cycle, however functional differences cannot be attributed to cell cycle progression alone.

### 3.5 Cfp1 is critical to regulation of gene expression

Previously, we identified altered H3K4 trimethylation in ES cells lacking Cfp1 by confocal microscopy. The subnuclear distribution of H3K4 trimethyl specific fluorescent antibody colocalized with bright areas of the nucleus when stained with diamidino-2-phenylindole (DAPI), apparently spreading to regions of heterochromatin (Lee et al., 2007). Cfp1 null ES cells have increased levels of Dnmt1 transcripts, but lower levels of Dnmt1 protein, and lower levels of protein in general (Butler et al., 2009). Upon induction of differentiation, Cfp1 deficient ES cells failed to differentiate and down-regulate the pluripotent stem cell marker Oct4 (Tate et al., 2009b). Altogether, the data suggests aberrant gene regulation and expression in absence of Cfp1. In order to better understand which genes are affected by the absence of Cfp1, microarray analysis was performed using two independent Cfp1 null ES cell lines generated from the same blastocyst harvest designated as N1 and N2. The null ES cells were compared to WT cells and an N1 cell line transfected with human Cfp1 cDNA (Rescue) which rescued phenotypic defects observed in CFP1 null cells and expresses approximate WT levels of Cfp1.

Initial analysis of the microarray data showed tight clustering of the four biological replicates within each cell line by both Principal Component Analysis (PCA) and hierarchical clustering indicating reliable identity within the cell lines (Figure 8). The WT and Rescue clones clustered tightly together as a group and separated distinctly from N1 and N2 most obviously by the PCA component 1 (x-axis) with 18.6% variance indicating commonality of WT and Rescue compared to Nulls as a group. However, N1 and N2 displayed somewhat divergent characteristics from each other in the PCA

analysis observed in y-axis PCA component 2 with 12.9% variance. Hierarchical clustering grouped the two nulls together with only a slightly larger distance between them compared to the distance between WT and Rescue validating distinct groupings between cell lines with and without Cfp1.

A one-way analysis of variation (ANOVA) was performed using the top 100 probes by p-value (Figure 9) to evaluate expression changes over all four cell lines. A comparison of N1, WT and Rescue showed that a majority of mis-regulated genes were rescued, however many of the genes were not. Of those not rescued, there were a larger number of up-regulated genes. Notably, many of these were also not up-regulated in the N2 cell line. Identification of both similarities and differences between the two Nulls suggests some but not all mis-regulated genes are directly related to the absence of Cfp1.

To better identify consistency in difference and similarities between WT and Cfp1 null, a two-way ANOVA was performed. The average of WT and Rescue was designated as "WT genotype" and the average of N1 and N2 was designated as "Null genotype". The two-way ANOVA compares the samples using both genotype and individual cell line. Hierarchical clustering of the top 100 by p-value using two-way ANOVA identified more genes consistently up and down regulated across the two "genotypes" compared to one-way ANOVA clustering (Figure 10), demonstrating increased utility for identification of Cfp1 affected genes. FDR (q-value), p-value and fold changes were calculated for ES cell lines using both genotype and individual cell line calculations to make contrasts between the cell types (data not shown). Two-way ANOVA analysis reveals a large set of statistical differences in contrasts at $p < .05$, with 12560 in N1 vs. WT (48%), 11582 in N1 vs. Rescue (44%), 9303 in Rescue vs. WT

(39%), 10260 in N2 vs. N1 (39%) and 12044 in N2 vs. WT (46%) (Table 2). Even at p< .0001, the number of differences identified in the contrasts remained high with 4471 in N1 vs. WT (17%), 3537 in N1 vs. Rescue (14%), 1927 in Rescue vs. WT (7%), 2741 in N2 vs. N1 (11%) and 3777 in N2 vs. WT (14%) indicating loss of Cfp1 dramatically alters gene expression.

Genes consistently up-regulated or down-regulated in both N1 and N2 which are rescued by re-introduction of Cfp1 should indicate a direct correlation to Cfp1. Two-way ANOVA contrasts were evaluated for consistency in directional mis-regulation using N1 vs. WT, N2 vs. WT followed by N1 vs. Rescue contrasts. Again, large numbers of probes were identified with approximately 29% consistently up or down-regulated in N1 and N2 ($p<.05$) and 18% rescued (Table 3). Consistent with top 100 by p-value in one-way ANOVA, more up-regulated genes did not rescue compared to down-regulated genes. However less than 1% of the total genes consistently mis-regulated were not rescued when using fold change of +/-1.5 as a filter. The inability to rescue genes may indicate changes in DNA methylation or other epigenetic marks that prevent Cfp1 association with previous targets. Altogether, this data suggests Cfp1 is a critical regulator of both activation and repression of gene expression.

### 3.6 Multiple cellular processes are impacted by loss of Cfp1

In order to better understand the biological role of genes affected by the absence of Cfp1, the large probe lists were minimized by using highly significant p-value ($p<.0001$) and consistent fold changes in N1, N2 and Rescue vs. WT of +/- 1.5 in two-way ANOVA contrasts. Using this filter, 686 probes representing 569 distinct genes

were identified, with 60% (343) up-regulated and 40% down-regulated (226) (Tables 4 and 5, respectively). The up and down-regulated gene lists were submitted to the Database for Annotation, Visualization, Identification and Discovery (DAVID) (http://david.abcc.ncifcrf.gov/), a web-based functional annotation analysis tool which accesses multiple databases in a single query. The Gene Ontology Consortium (GO) database enables up-to-date controlled vocabularies to describe the role of genes and gene products with three categories including molecular function (GO_MF), biological process (GO_BP) and cellular component (GO_CC). Terms within each of these categories are structured to allow a broad range of inquiry, from generic classification to detailed terms within the classifications. Gene lists submitted are evaluated for enrichment within categories.

GO_BP analysis associated over 50% of up and down regulated genes with the general term *cellular process*. Overall, 138 different GO_BP terms were associated with the down-regulated genes. The second highest category associated with down-regulated genes was *primary metabolic process* making up 37% (87) of the list, followed by 21% (49) associated with *nucleobase, nucleoside, nucleotide and nucleic acid metabolic process* and 18% (42) associated with *regulation of process* (Table 6). Up-regulated genes were associated with 171 GO_BP terms overall. The second highest category comprised 35% (120) of up-regulated genes and associated with *biological regulation*, followed by 14% (49) associated with *multicellular organismal development*, and 14% (48) genes associated with *gene expression* (Table 7).

The Protein Analyis Through Evolutionary Relationships (PANTHER) classification system classifies genes by their function. PANTHER molecular function

analysis enables a more meaningful analysis of molecular function by eliminating top tier GO terms for molecular function. For example, "receptor" or "transcription factor" is more meaningful than "protein binding" or "transcription factor binding". Sixteen PANTHER molecular functions were associated with up-regulated genes and nine associated with down-regulated genes. *Nucleic acid binding* topped both lists with 20% (47) of down-regulated and 11% (38) of up regulated genes (Table 8). The second highest molecular function identified in each list differed, with 17% (39) down-regulated genes associated with *transcription factor* in contrast to 4% (14) up-regulated genes associated with *hydrolase*.

The Kyoto Encyclopedia of Genes and Genomes (Kegg) Pathway is a collection of manually drawn pathway maps based on the known molecular interactions for metabolism, cellular process, human diseases, organismal systems and genetic and environmental information processing (http://www. genome.ad.jp/kegg/) (Kanehisa & Goto, 2000). Both up and down-regulated genes were enriched in the *glycolysis/gluconeogenesis* pathway (Table 9). Down regulated genes were enriched in *aminoacyl-tRNA biosynthesis*, *spliceosome*, *TGF-beta signaling*, *ribosome,* and *Melanoma* pathways. Up-regulated genes were enriched in *Alzheimer's disease*, *ether lipid metabolism*, *complement and coagulation cascades*, *glycerolipid metabolism*, *glycerophospholipid metabolism* and *tyrosine metabolism*. Taken together, these data suggest Cfp1 plays a critical role in regulation of basic cellular processes and signaling in addition to development and gene expression.

In order to identify the biological role of genes which did not rescue, a gene list was generated with less stringent parameters due to the smaller data set. Utilizing

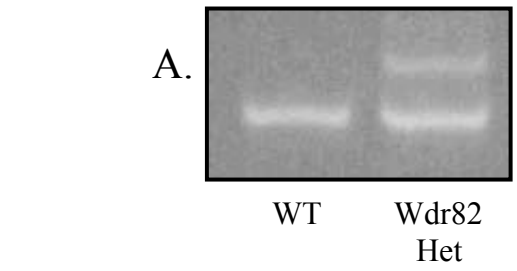
consistency in up or down-regulation with disregard to fold change levels and $p<.05$, only 138 down regulated and 178 up-regulated genes were not rescued.

GO_BP associations with non-rescued up-regulated genes were *regulation of apoptosis* (13), *RNA processing* (12) and *macromolecular catabolic process* (11). PANTHER molecular function analysis associated the list with *glycosyltransferase, actin binding* and *cytoskelatal protein*. Kegg pathways identified were *spliceosome, Fc gamma R-mediated phagocytosis, N-glycan biosynthesis* and *long term depression*. Further filtering with the fold change of 1.5 reduced the list to 78 genes, however the associated biological processes identified were consistent (Table 10). PANTHER molecular function only identified *receptor* and no Kegg pathways were flagged (data not shown).

Non-rescued down-regulated genes were associated with *cell cycle*, *mitotic cell cycle* and *cell division* in GO_BP analysis as primary processes. PANTHER molecular function associated 21 genes with *nucleic acid binding*, and others were associated with *methyltransferases, hydrolase* and *"other" phosphatase*. *Oocyte meiosis* and *ubiquitin mediated proteolysis* were the only two Kegg pathways implicated. Further filtering with a 1.5 cutoff reduced this list to only 38 genes and *cell cycle* was consistently identified as a primary GO_BP term (Table 11). However, *anatomical structure development* and *enzyme linked receptor* were also terms identified. No PANTHER molecular function or Kegg Pathways were associated with the fold change filtered list.

The summary of biological function analysis reveal that Cfp1 is involved in multiple critical cellular processes, including cell cycle control, development and

biosynthetic and metabolic processes, some of which may not be rescued by Cfp1 replacement.



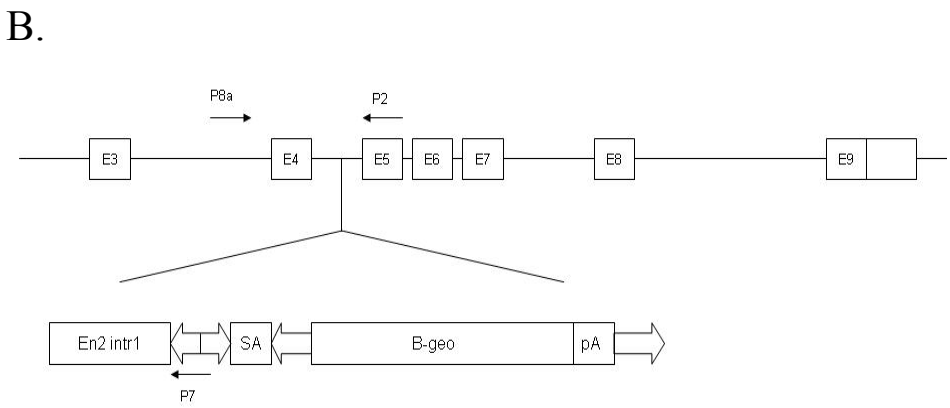




**Figure 1. PCR products from WT and Wdr82$^{(+/-)}$ embryo DNA.**
A. Embryo DNA was analyzed for genotype by multiplex PCR products run on 1% agarose gel and identified by ethidium bromide. Size was determined by standard 100 base pair DNA ladder. B. Cartoon depicting Wdr82 gene location for primers used in multiplex PCR. B-geo is the LacZ gene casette coding for β-galactosidase.

**Table 1. Embryo genotypes in Wdr82$^{(+/-)}$ matings.**
Wdr82$^{(+/-)}$ mice were mated and embryos collected on day 9.5–11.5 dpc for DNA extraction. Genotype analysis was performed by multiplex PCR amplification as described in materials and methods. Amplification products were detected with ethidium bromide and run on 1.5% agarose gel.

| Time point | Total Embryos | WT | Wdr82 Het | Wdr82 Nulls |
|---|---|---|---|---|
| 9.5-11.5 dpc | 74 | 7 | 67 | 0 |

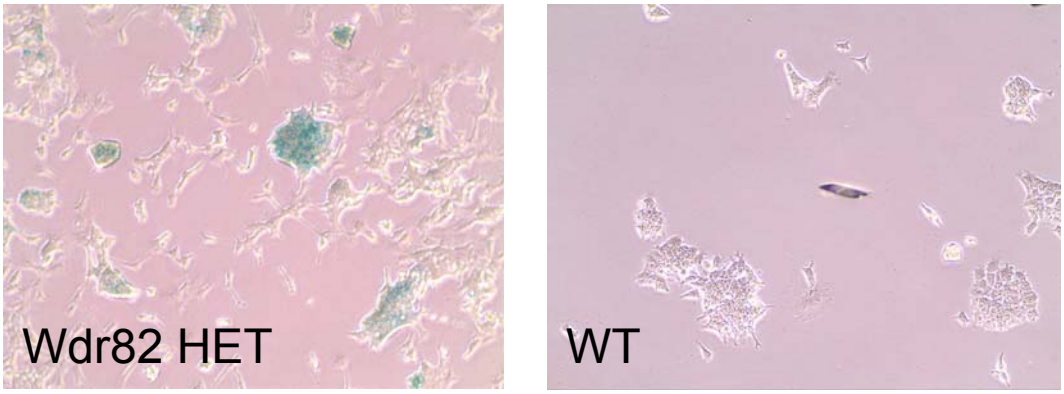



**Figure 2. Wdr82 Het ES cells demonstrate β-galactosidase activity.**
Asynchronous exponentially growing Wild Type (WT) and Wdr82 Het ES cells containing the LacZ disrupted gene trap (AZ0280) were stained for β-galactosidase activity in X-gal staining buffer. Cells were observed under inverted light microscope at 10x magnification. β-galactosidase activity is indicated by presence of blue color.

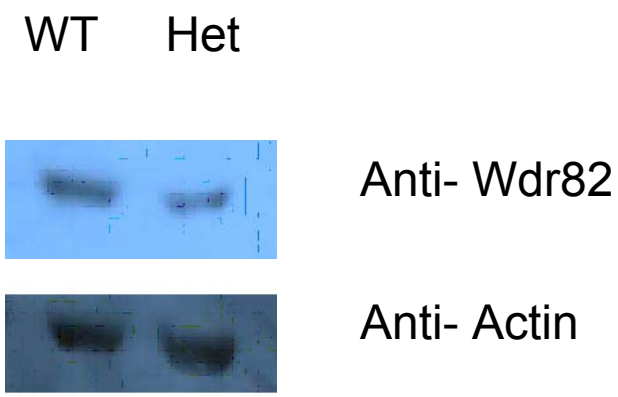



**Figure 3. Wdr82 het ES cell produce reduced levels of Wdr82 protein.**
Nuclear enriched cell lysates collected from asynchronous exponentially growing WT and Wdr82 Het ES cells grown in LIF to prevent differentiation were analyzed by Western blot with antibodies indicated. β-actin demonstrates relative protein levels in sample.

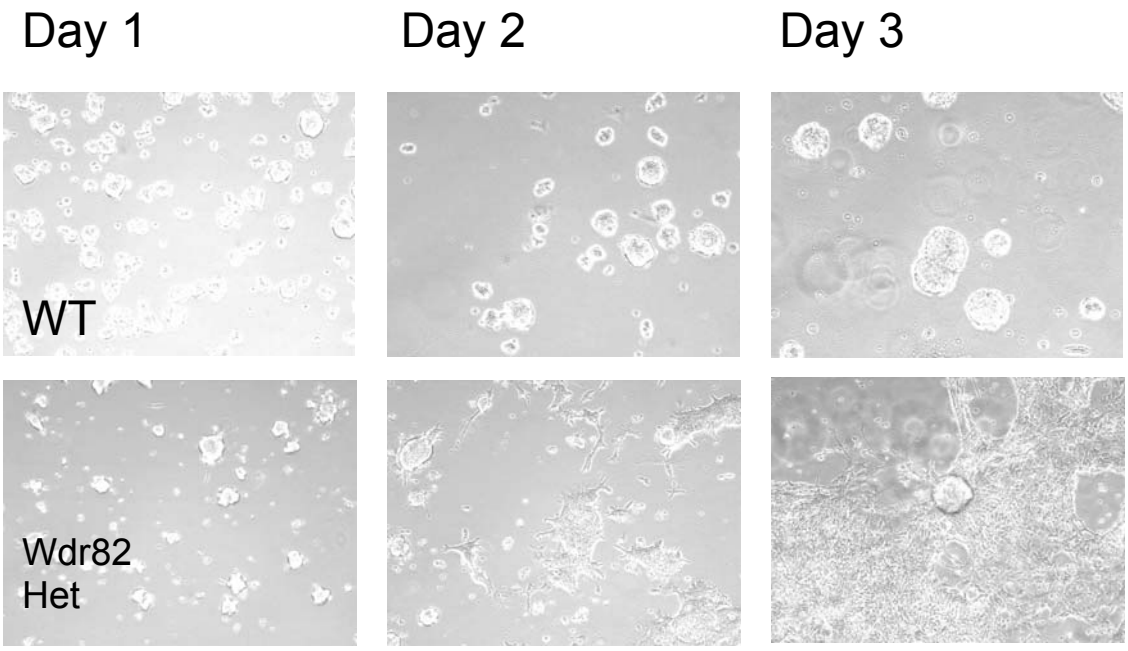



**Figure 4. Wdr82 Het cells exhibit earlier induction of differentiation.**
WT and Wdr82 Het ES cells were plated in ES culture media without LIF to induce differentiation. Cells were monitored by light microscope in 24 hour intervals. Outgrowth of embroid-like bodies is indicative of differentiation.

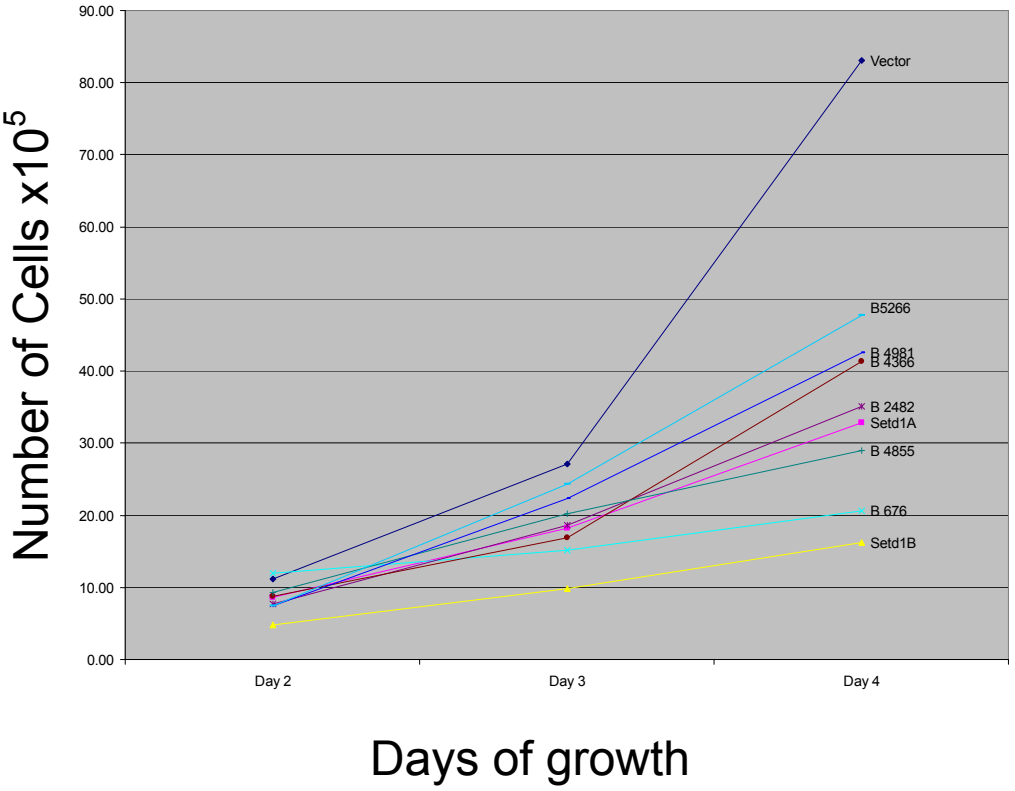



**Figure 5. Cell counts for cDNA expressing HEK 293 T-rex cells.**
HEK 293 T-rex cells were seeded in duplicate and expression induced by addition of 1 mg/ml doxycycline. Cells were collected at day 1, 2, 3 and 4 in 24 hour intervals. Day 2, 3 and 4 cell counts are shown above. Viable cells were counted by hemocytometer with Trypan Blue staining under an inverted light microscope.

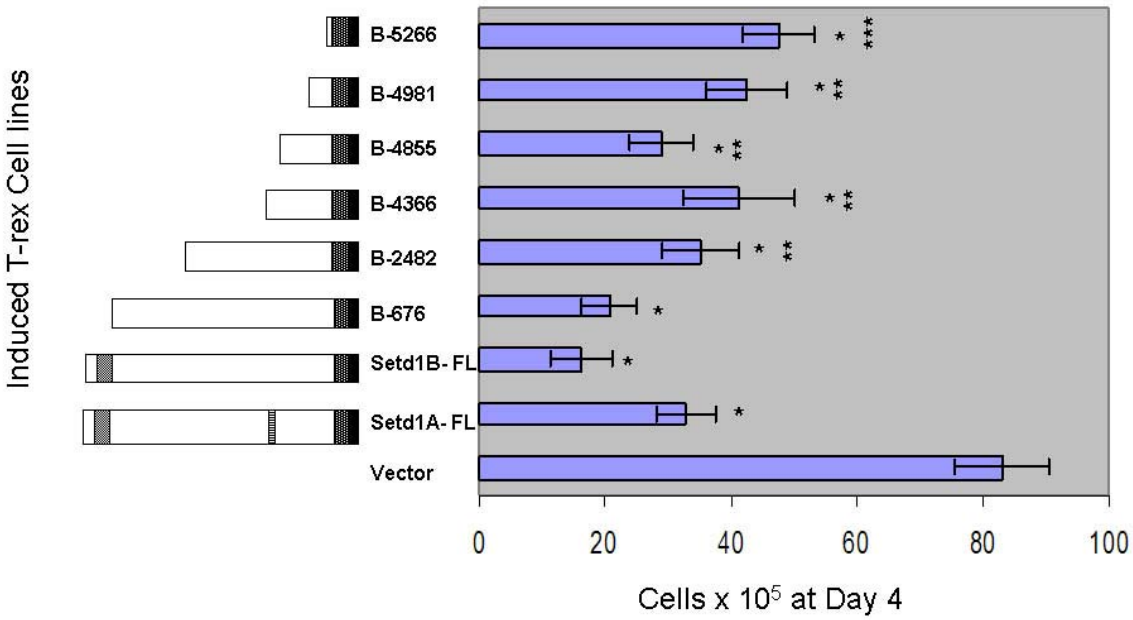



**Figure 6. Significantly reduced cell counts at day 4 with overexpression of Setd1 full length or truncated cDNA.**
HEK 293 T-rex cells were seeded in duplicate and expression induced by addition of 1 mg/ml doxycycline. Statistical analysis was performed on cell counts for day 4 shown above. Viable cells were counted by hemocytometer with Trypan Blue staining under inverted light microscope. * Significant difference ($p<0.0001$) than vector. ** Significant difference ($p<0.005$) than Setd1B, but not Setd1A. *** Significant difference ($p<0.005$) than Setd1A, Setd1B and vector.

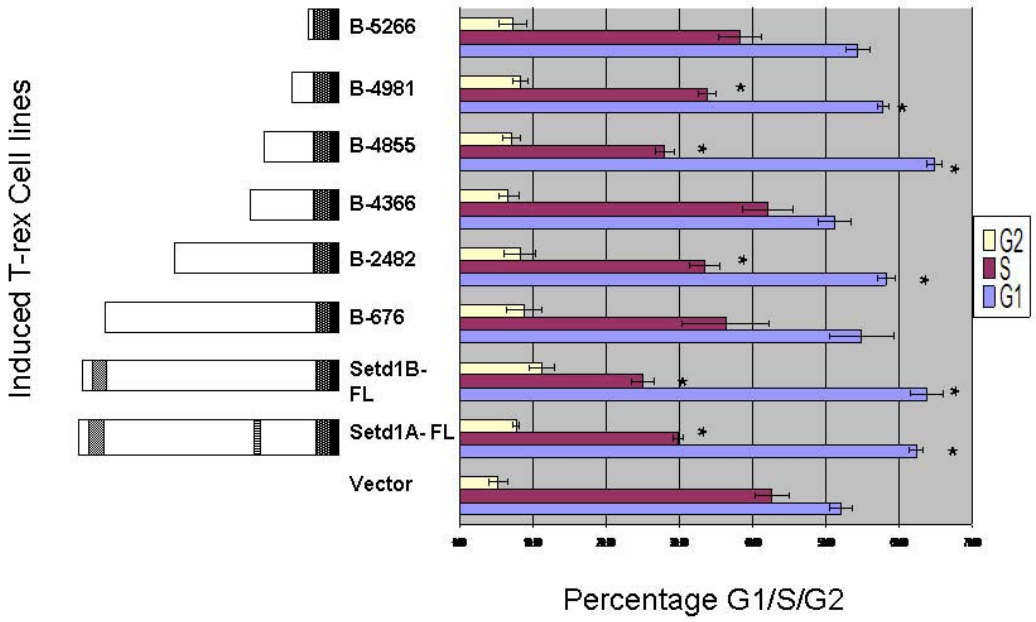
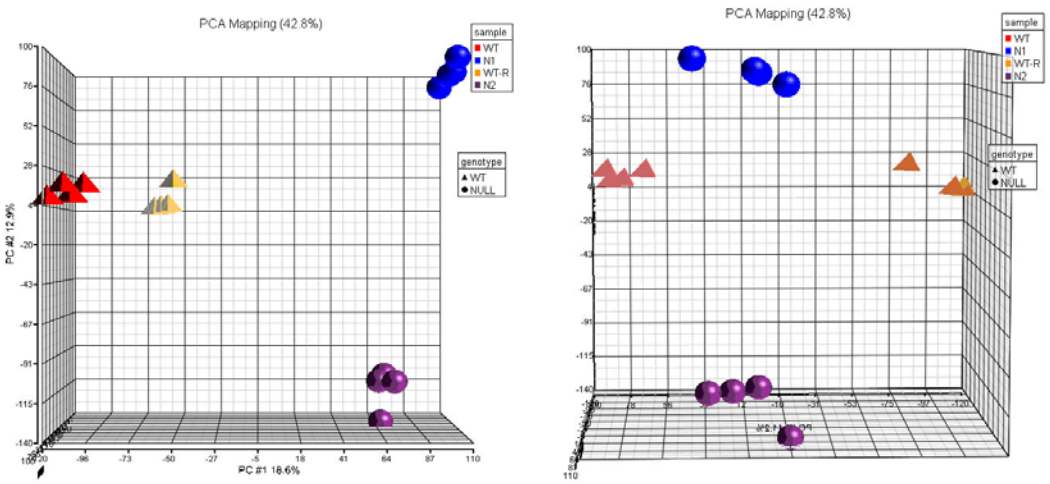



**Figure 7. Setd1 overexpression alters cell cycle.**
Stably transfected inducible HEK 293 T-rex cells were grown in media with doxycycline to induce gene expression. Asynchronous exponentially growing cells were collected and analyzed for cell cycle by propidium iodide staining on FACScan flow cytometer as described in methods and materials. * Statisitical significance $p<0.01$ compared to vector.

A.




B.

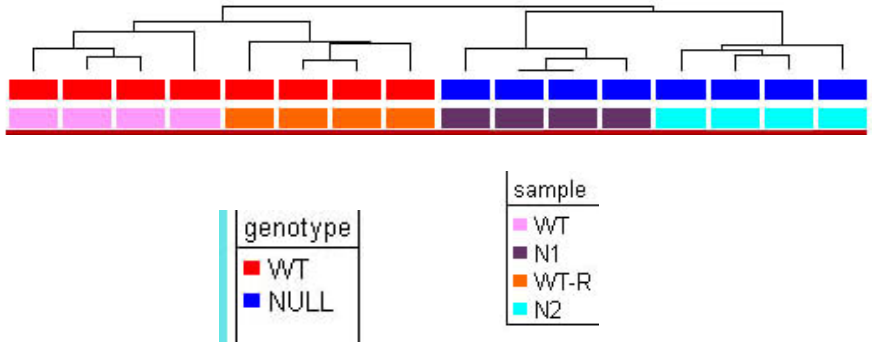



**Figure 8. CFP1$^{(-/-)}$ ES cell microarray data tightly cluster and separate.**
Total RNA from one wild type (WT), two distinct Cfp1$^{(-/-)}$ (N1 and N2) and a Cfp1 knockout cell line (N1) transfected with human Cfp1 cDNA (WT-R) was analyzed by Affymetrix Mouse Genome 430 2.0 GeneChip®. Overall Data was log transformed (base 2) in Park Genomics Suite software for quality assessment. A. Principal Component Analysis in three dimensions. B. Heirachechal clustering of the same overall data set.

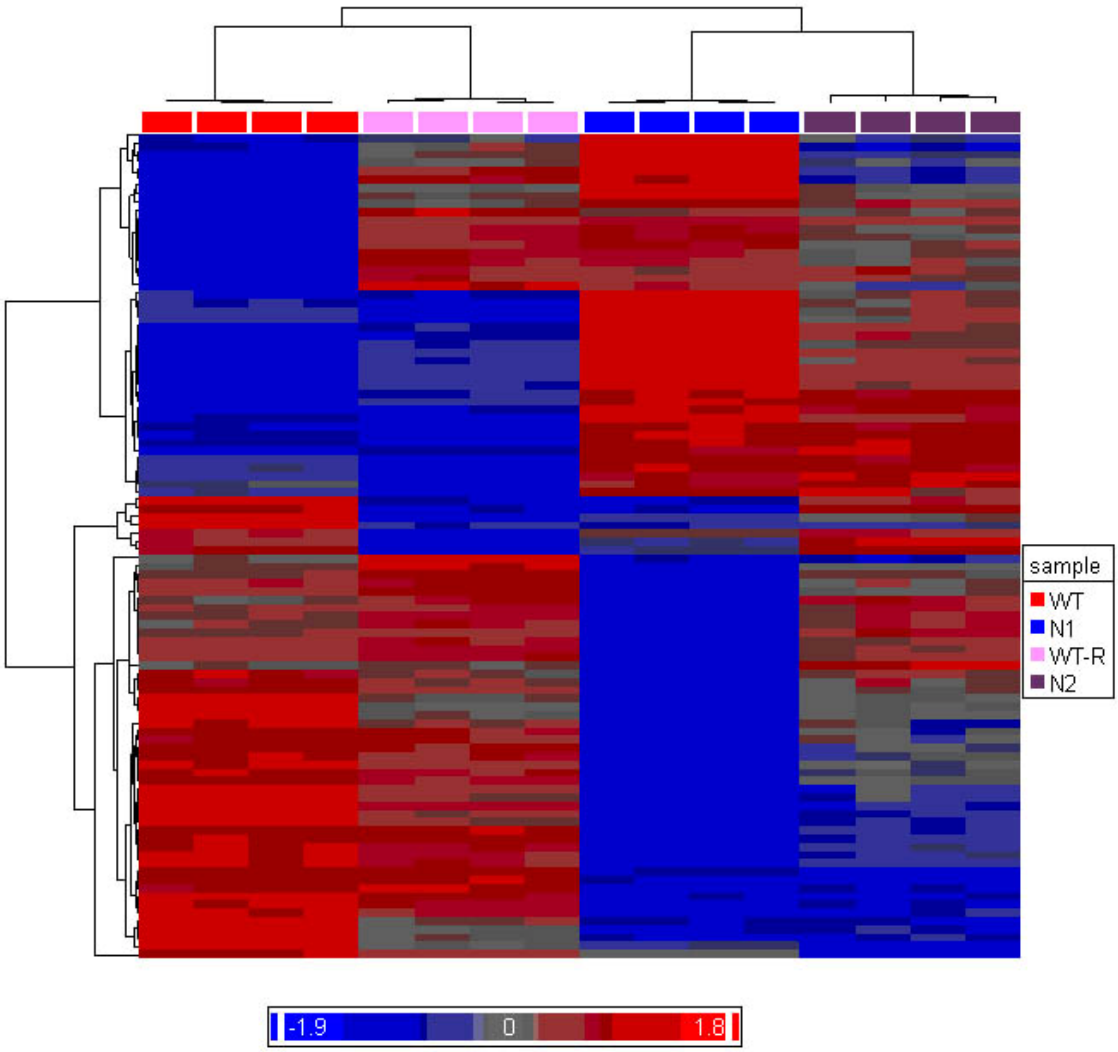



**Figure 9. Top 100 changes in gene expression identified by p-value with one-way analysis of variation.**
Total RNA from one wild type (WT), two distinct CFP1 knockout (N1 and N2) and a CFP1 knockout cell line (N1) transfected with human CFP1cDNA (WT-R) was analyzed by Affymetrix Mouse Genome 430 2.0 GeneChip®. Color scale indicates relative expression levels after log transformation and normalization of data as described in materials and methods in one-way analysis of variation.

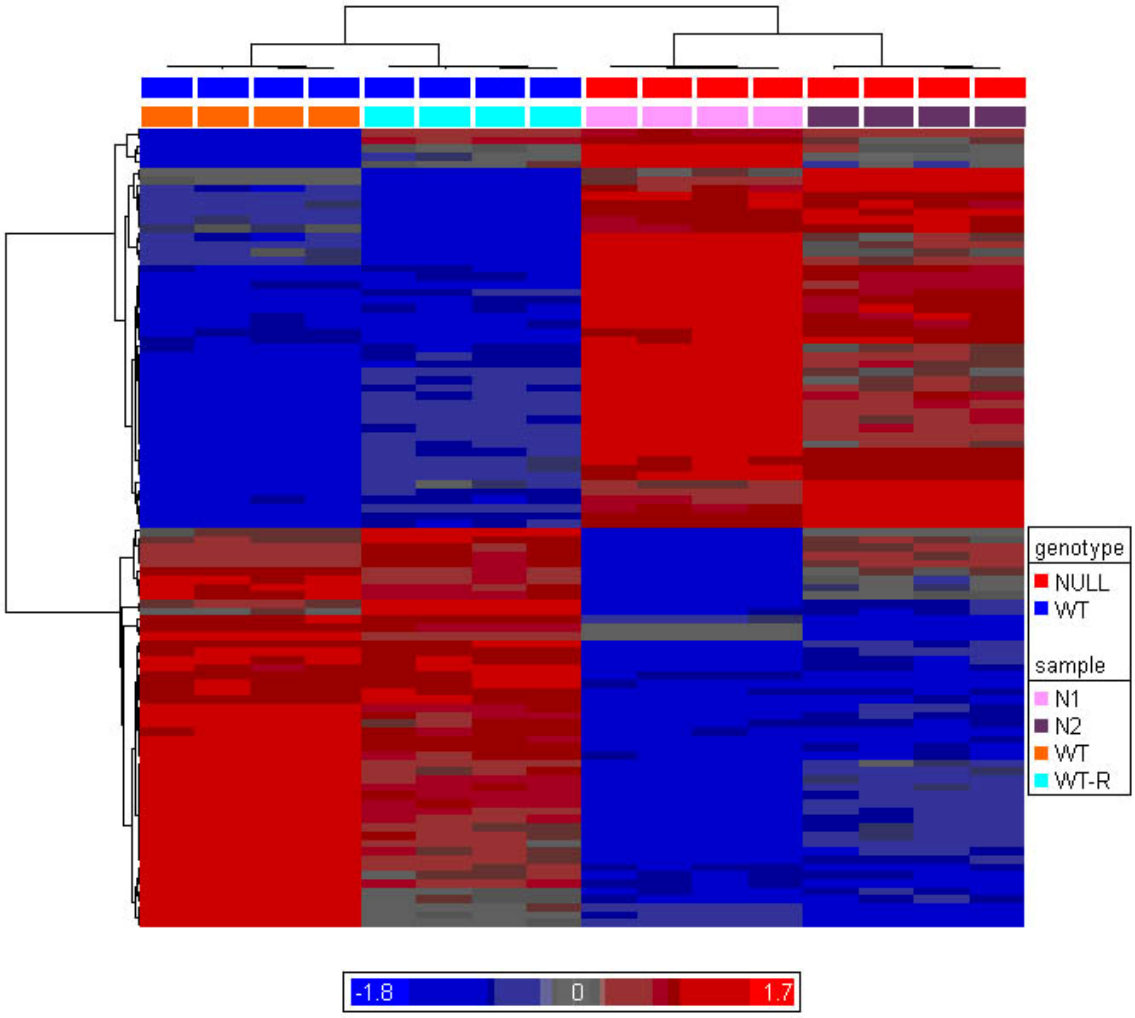



**Figure 10. Top 100 changes in gene expression identified by p-value with two-way analysis of variation.**
Total RNA from one wild type (WT), two distinct Cfp1 knockout (N1 and N2) and a Cfp1 knockout cell line (N1) transfected with human Cfp1 cDNA (WT-R) was analyzed by Affymetrix Mouse Genome 430 2.0 GeneChip®. Color scale indicates relative expression levels after log transformation and normalization of data as described in materials and methods in two-way analysis of variation.

**Table 2. Number of probes identified as statistically different in two-way ANOVA contrasts.**

| | <.0001 | <0.001 | <0.01 | <0.025 | <0.05 | <0.1 | <1 |
|---|---|---|---|---|---|---|---|
| **NULL * N1 vs WT * WT (N1 vs WT)** | | | | | | | |
| p-value | 4471 | 6481 | 9441 | 11041 | 12560 | 14363 | 26082 |
| q-value (FDR) | 3782 | 6034 | 9403 | 11417 | 13350 | 16012 | 26085 |
| | | | | | | | |
| **NULL * N2 vs WT * WT (N2 vs WT)** | | | | | | | |
| p-value | 3777 | 5839 | 8824 | 10507 | 12044 | 13918 | 26078 |
| q-value (FDR) | 3009 | 5191 | 8505 | 10621 | 12596 | 15256 | 26085 |
| | | | | | | | |
| **NULL * N2 vs NULL * N1 (N1 vs N2)** | | | | | | | |
| p-value | 2741 | 4361 | 7091 | 8649 | 10260 | 12290 | 26080 |
| q-value (FDR) | 1866 | 3416 | 6216 | 7942 | 9854 | 12592 | 26085 |
| | | | | | | | |
| **WT * WT-R vs WT * WT (Rescue vs WT)** | | | | | | | |
| p-value | 1927 | 3367 | 6035 | 7726 | 9303 | 11333 | 26080 |
| q-value (FDR) | 1030 | 2199 | 4615 | 6388 | 8330 | 10825 | 26085 |
| | | | | | | | |
| **NULL * N1 vs WT * WT-R (N1 vs Rescue)** | | | | | | | |
| p-value | 3537 | 5469 | 8431 | 10008 | 11582 | 13524 | 26082 |
| q-value (FDR) | 2717 | 4681 | 8068 | 9969 | 12005 | 14765 | 26085 |

**Table 3. Number of probes identified as up- and down-regulated by p-value.**
Two-way ANOVA contrasts were analyzed for consistency in up and down regulation in N1 and N2 vs. WT and N1 vs. Rescue as described in materials and methods.

| | No filter | p<.05 | % Present | p<.01 | % Present | p<.001 | % Present | p<.0001 | % Present |
|---|---|---|---|---|---|---|---|---|---|
| N1 vs WT <0 | 12889 | 6156 | 23.6 | 4570 | 17.5 | 3140 | 12.0 | 2132 | 8.2 |
| N1 vs WT >0 | 13196 | 6410 | 24.6 | 4872 | 18.7 | 3341 | 12.8 | 2339 | 9.0 |
| **Total** | **26085** | **12566** | **48.2** | **9442** | **36.2** | **6481** | 24.8 | **4471** | **17.1** |
| | | | | | | | | | |
| N2 vs WT <0 | 11939 | 5224 | 20.0 | 3840 | 14.7 | 2540 | 9.7 | 1662 | 6.4 |
| N2 vs WT >0 | 14146 | 6820 | 26.1 | 4984 | 19.1 | 3299 | 12.6 | 2115 | 8.1 |
| **Total** | **26085** | **12044** | **46.2** | **8824** | **33.8** | **5839** | 22.4 | **3777** | **14.5** |
| | | | | | | | | | |
| N1 & N2 vs WT <0 | 8876 | 3307 | 12.7 | 2345 | 9.0 | 1523 | 5.8 | 958 | 3.7 |
| N1 & N2 vs WT>0 | 10133 | 4184 | 16.0 | 2972 | 11.4 | 1893 | 7.3 | 1229 | 4.7 |
| **Total** | **19009** | **7491** | **28.7** | **5317** | **20.4** | **3416** | 13.1 | **2187** | **8.4** |
| | | | | | | | | | |
| N1 & N2 vs WT <0 = Rescue | 6936 | 2102 | 8.1 | 1472 | 5.6 | 888 | 3.4 | 537 | 2.1 |
| N1 & N2 vs WT >0 =Rescue | 7123 | 2542 | 9.7 | 1737 | 6.7 | 1050 | 4.0 | 653 | 2.5 |
| **Total** | **14059** | **4644** | **17.8** | **3209** | **12.3** | **1938** | 7.4 | **1190** | **4.6** |
| | | | | | | | | | |
| N1 & N2 vs WT <0; No Rescue | 1940 | 138 | 0.5 | 58 | 0.2 | 29 | 0.1 | 9 | 0.0 |
| N1 & N2 vs WT >0; No Rescue | 3010 | 178 | 0.7 | 72 | 0.3 | 52 | 0.2 | 7 | 0.0 |
| **Total** | **4950** | **316** | **1.2** | **130** | **0.5** | **81** | 0.3 | **16** | 0.1 |

**Table 4. Up-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis.**

Genes were identified using two-way ANOVA contrasts N1 vs. WT, N1 vs. Rescue and N2 vs. WT ($p<0.0001$) with fold change ≥ 1.5.

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 1** | | **Chromosome 3** | |
| **Vnn1** | vanin 1 | **3200002M19Rik** | RIKEN cDNA 3200002M19 gene |
| **2310035C23Rik** | RIKEN cDNA 2310035C23 gene | **5730508B09Rik** | RIKEN cDNA 5730508B09 gene |
| **Agfg1** | ArfGAP with FG repeats 1 | **Adh1** | alcohol dehydrogenase 1 (class I) |
| **Aox4** | aldehyde oxidase 4 | **Bhlhe22** | basic helix-loop-helix family, member e22 |
| **C130073F10Rik** | RIKEN cDNA C130073F10 gene | **Crabp2** | cellular retinoic acid binding protein II |
| **Cd55** | CD55 antigen | **Fam198b** | family with sequence similarity 198, member B |
| **Cnst** | consortin, connexin sorting protein | **Fndc3b** | fibronectin type III domain containing 3B |
| **D630023F18Rik** | RIKEN cDNA D630023F18 gene | **Igsf3** | immunoglobulin superfamily, member 3 |
| **Ddr2** | discoidin domain receptor family, member 2 | **Mfsd1** | major facilitator superfamily domain containing 1 |
| **Gpa33** | glycoprotein A33 (transmembrane) | **Mme** | membrane metallo endopeptidase |
| **Igfbp2** | insulin-like growth factor binding protein 2 | **Ppp3ca** | protein phosphatase 3, catalytic subunit, alpha isoform |
| **LOC100044384** | hypothetical protein LOC100044384 | **Ptpn22** | protein tyrosine phosphatase, non-receptor type 22 (lymphoid) |
| **Msc** | musculin | **Snx16** | sorting nexin 16 |
| **Ncoa2** | nuclear receptor coactivator 2 | **Sohlh2** | spermatogenesis and oogenesis specific basic helix-loop-helix 2 |
| **Pign** | phosphatidylinositol glycan anchor biosynthesis, class N | **Tlr2** | toll-like receptor 2 |
| **Prdm14** | PR domain containing 14 | **Zbtb10** | zinc finger and BTB domain containing 10 |
| **Ren1 /// Ren2** | renin 1 structural /// renin 2 tandem duplication of Ren1 | **Chromosome 4** | |
| **Satb2** | special AT-rich sequence binding protein 2 | **Aqp3** | aquaporin 3 |
| **Stk17b** | serine/threonine kinase 17b (apoptosis-inducing) | **Btbd19** | BTB (POZ) domain containing 19 |
| **Chromosome 2** | | **Car8** | carbonic anhydrase 8 |
| **Bmi1** | Bmi1 polycomb ring finger oncogene | **Ckb** | creatine kinase, brain |
| **Commd3** | COMM domain containing 3 | **Dctn3** | dynactin 3 |
| **Cst3** | cystatin C | **Fabp3** | fatty acid binding protein 3, muscle and heart |
| **Dpp4** | dipeptidylpeptidase 4 | **Grhl3** | grainyhead-like 3 (Drosophila) |
| **Meis2** | Meis homeobox 2 | **Lpar1** | lysophosphatidic acid receptor 1 |
| **Pramel6** | preferentially expressed antigen in melanoma like 6 | **Manea** | mannosidase, endo-alpha |
| **Ptges** | prostaglandin E synthase | **Ppap2b** | phosphatidic acid phosphatase type 2B |
| **Ryr3** | ryanodine receptor 3 | **Rpl22** | ribosomal protein L22 |
| **Serping1** | serine (or cysteine) peptidase inhibitor, clade G, member 1 | **Tek** | endothelial-specific receptor tyrosine kinase |
| **Smtnl1** | smoothelin-like 1 | **Chromosome 5** | |
| **Spcs2** | signal peptidase complex subunit 2 homolog (S. cerevisiae) | **Abcb1b** | ATP-binding cassette, sub-family B (MDR/TAP), member 1B |
| **Tfpi** | tissue factor pathway inhibitor | **Agpat9** | 1-acylglycerol-3-phosphate O-acyltransferase 9 |
| **Tnfaip6** | tumor necrosis factor alpha induced protein 6 | **Atp8a1** | ATPase, aminophospholipid transporter (APLT), class I, type 8A, member 1 |
| **Ube2l6** | ubiquitin-conjugating enzyme E2L 6 | **Cox7b2** | cytochrome c oxidase subunit VIIb2 |
| **Wfdc15a** | WAP four-disulfide core domain 15A | **Cpeb2** | cytoplasmic polyadenylation element binding protein 2 |

## Table 4. Up-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis. (cont.)

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 5 (cont.)** | | **Chromosome 7 (cont.)** | |
| **Limch1** | LIM and calponin homology domains 1 | **Btbd10** | BTB (POZ) domain containing 10 |
| **Lrpap1** | low density lipoprotein receptor-related protein associated protein 1 | **Bub3** | budding uninhibited by benzimidazoles 3 homolog (S. cerevisiae) |
| **Naa11** | N(alpha)-acetyltransferase 11, NatA catalytic subunit | **C2cd3** | C2 calcium-dependent domain containing 3 |
| **Otop1** | otopetrin 1 | **Calm3** | calmodulin 3 |
| **Pgm1** | phosphoglucomutase 1 | **Ceacam20** | carcinoembryonic antigen-related cell adhesion molecule 20 |
| **Slc5a1** | solute carrier family 5 (sodium/glucose cotransporter), member 1 | **Cib1** | calcium and integrin binding 1 (calmyrin) |
| **Tbx3** | T-box 3 | **Ctr9** | Ctr9, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae) |
| **Ulk1** | Unc-51 like kinase 1 (C. elegans) | **Cttn** | cortactin |
| **Chromosome 6** | | **Dcun1d3** | DCN1, defective in cullin neddylation 1, domain containing 3 (S. cerevisiae) |
| **A2m** | alpha-2-macroglobulin | **Dhx32** | DEAH (Asp-Glu-Ala-His) box polypeptide 32 |
| **Aebp2** | AE binding protein 2 | **Dmwd** | dystrophia myotonica-containing WD repeat motif |
| **Asz1** | ankyrin repeat, SAM and basic leucine zipper domain containing 1 | **Fam169b** | family with sequence similarity 169, member B |
| **Eps8** | epidermal growth factor receptor pathway substrate 8 | **Fam175b** | family with sequence similarity 175, member B |
| **Hipk2** | homeodomain interacting protein kinase 2 | **Far1** | fatty acyl CoA reductase 1 |
| **Ldhb** | lactate dehydrogenase B | **Fgf4** | fibroblast growth factor 4 |
| **Rimklb** | ribosomal modification protein rimK-like family member B | **Fgfr2** | fibroblast growth factor receptor 2 |
| **Mdfic** | MyoD family inhibitor domain containing | **Folr1** | folate receptor 1 (adult) |
| **Mfap5** | microfibrillar associated protein 5 | **Folr2** | folate receptor 2 (fetal) |
| **Pzp** | pregnancy zone protein | **Fxyd5** | FXYD domain-containing ion transport regulator 5 |
| **Slc35b4** | solute carrier family 35, member B4 | **Gde1** | glycerophosphodiester phosphodiesterase 1 |
| **Ube2h** | ubiquitin-conjugating enzyme E2H | **Polr2l /// Usp50** | polymerase (RNA) II (DNA directed) polypeptide L /// ubiquitin specific peptidase 50 |
| **Zc3hav1** | zinc finger CCCH type, antiviral 1 | **Gramd1a** | GRAM domain containing 1A |
| **Zxdc** | ZXD family zinc finger C | **H19** | H19 fetal liver mRNA |
| **Chromosome 7** | | **H47** | histocompatibility 47 |
| **2210018M11Rik** | RIKEN cDNA 2210018M11 gene | **Hdgfrp3** | hepatoma-derived growth factor, related protein 3 |
| **LOC100047683** | hypothetical protein LOC100047683 | **Hps5** | Hermansky-Pudlak syndrome 5 homolog (human) |
| **6330503K22Rik** | RIKEN cDNA 6330503K22 gene | **Htatip2** | HIV-1 tat interactive protein 2, homolog (human) |
| **9030624J02Rik** | RIKEN cDNA 9030624J02 gene | **Htra1** | HtrA serine peptidase 1 |
| **Alg8** | asparagine-linked glycosylation 8 homolog (yeast, alpha-1,3-glucosyltransferase) | **Idh2** | isocitrate dehydrogenase 2 (NADP+), mitochondrial |
| **Arrdc4** | arrestin domain containing 4 | **Igf1r** | insulin-like growth factor I receptor |
| **Ate1** | arginyltransferase 1 | **Kctd15** | potassium channel tetramerisation domain containing 15 |
| **AU018091** | expressed sequence AU018091 | **Kndc1** | kinase non-catalytic C-lobe domain (KIND) containing 1 |
| **Aurkc** | aurora kinase C | **Ldhc** | lactate dehydrogenase C |
| **B230311B06Rik** | RIKEN cDNA B230311B06 gene | **Lrrc28** | leucine rich repeat containing 28 |
| **BC066028** | cDNA sequence BC066028 | **Lsm14a** | LSM14 homolog A (SCD6, S. cerevisiae) |
| **Bola2** | bolA-like 2 (E. coli) | **Mex3b** | mex3 homolog B (C. elegans) |

## Table 4. Up-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis. (cont.)

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 7 (cont.)** | | **Chromosome 8 (cont.)** | |
| **Mfge8** | milk fat globule-EGF factor 8 protein | **Col4a2** | collagen, type IV, alpha 2 |
| **Mrpl23** | mitochondrial ribosomal protein L23 | **Dlc1** | deleted in liver cancer 1 |
| **Mrpl46** | mitochondrial ribosomal protein L46 | **Hgsnat** | heparan-alpha-glucosaminide N-acetyltransferase |
| **Mtmr10** | myotubularin related protein 10 | **Hpgd** | hydroxyprostaglandin dehydrogenase 15 (NAD) |
| **Ndufa3** | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3 | **Lpl** | lipoprotein lipase |
| **Ndufc2** | NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 2 | **Nqo1** | NAD(P)H dehydrogenase, quinone 1 |
| **Nfatc2ip** | nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein | **Slc35f3** | solute carrier family 35, member F3 |
| **Nlrp4c** | NLR family, pyrin domain containing 4C | **2410076I21Rik** | RIKEN cDNA 2410076I21 gene |
| **Pdcd2l** | programmed cell death 2-like | **9030425E11Rik** | RIKEN cDNA 9030425E11 gene |
| **Pih1d1** | PIH1 domain containing 1 | **Chromosome 9** | |
| **Pik3c2a** | phosphatidylinositol 3-kinase, C2 domain containing, alpha polypeptide | **BC031353** | cDNA sequence BC031353 |
| **Ppp1r14a** | protein phosphatase 1, regulatory (inhibitor) subunit 14A | **Ccpg1** | cell cycle progression 1 |
| **Prr14** | proline rich 14 | **Gm5918** | predicted gene 5918 |
| **Rab30** | RAB30, member RAS oncogene family | **Gsta1 /// Gsta2** | glutathione S-transferase, alpha 1 (Ya) /// glutathione S-transferase, alpha 2 (Yc2) |
| **Rabac1** | Rab acceptor 1 (prenylated) | **Lrrc2** | leucine rich repeat containing 2 |
| **Rbm42** | RNA binding motif protein 42 | **Rab27a** | RAB27A, member RAS oncogene family |
| **Rhpn2** | rhophilin, Rho GTPase binding protein 2 | **Smad6** | MAD homolog 6 (Drosophila) |
| **Rnf121** | ring finger protein 121 | **Spa17** | sperm autoantigenic protein 17 |
| **Rsf1** | remodeling and spacing factor 1 | **Spesp1** | sperm equatorial segment protein 1 |
| **Saa3** | serum amyloid A 3 | **Tmem41b** | transmembrane protein 41B |
| **Saal1** | serum amyloid A-like 1 | **Gsta1 /// Gsta2** | glutathione S-transferase, alpha 1 (Ya) /// glutathione S-transferase, alpha 2 (Yc2) |
| **Sars2** | seryl-aminoacyl-tRNA synthetase 2 | **Lrrc2** | leucine rich repeat containing 2 |
| **Siglec5** | sialic acid binding Ig-like lectin 5 | **Rab27a** | RAB27A, member RAS oncogene family |
| **Syce1** | synaptonemal complex central element protein 1 | **Smad6** | MAD homolog 6 (Drosophila) |
| **Tead1** | TEA domain family member 1 | **Spa17** | sperm autoantigenic protein 17 |
| **Tfpt** | TCF3 (E2A) fusion partner | **Spesp1** | sperm equatorial segment protein 1 |
| **Tmc7** | transmembrane channel-like gene family 7 | **Tmem41b** | transmembrane protein 41B |
| **Tnnt1** | troponin T1, skeletal, slow | **Chromosome 10** | |
| **Tubgcp2** | tubulin, gamma complex associated protein 2 | **AI317395** | expressed sequence AI317395 |
| **Wdr11** | WD repeat domain 11 | **Chpt1** | choline phosphotransferase 1 |
| **Wdr62** | WD repeat domain 62 | **Ipmk** | inositol polyphosphate multikinase |
| **Zbtb45** | zinc finger and BTB domain containing 45 | **Kitl** | kit ligand |
| **Zfp264** | zinc finger protein 264 | **LOC100047324 /// Sesn1** | similar to Sesn1 protein /// sestrin 1 |
| **Zfp689** | zinc finger protein 689 | **Map3k5** | mitogen-activated protein kinase kinase kinase 5 |
| **Zfp74** | zinc finger protein 74 | **Psap** | prosaposin |
| **Chromosome 8** | | **Rbms2** | RNA binding motif, single stranded interacting protein 2 |
| **Ccl25** | chemokine (C-C motif) ligand 25 | **Spic** | Spi-C transcription factor (Spi-1/PU.1 related) |
| **Col4a1** | collagen, type IV, alpha 1 | **Txlnb** | taxilin beta |

## Table 4. Up-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis. (cont.)

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 11** | | **Chromosome 11 (cont.)** | |
| **1500010J02Rik** | RIKEN cDNA 1500010J02 gene | **Sparc** | secreted acidic cysteine rich glycoprotein |
| **1700042O10Rik** | RIKEN cDNA 1700042O10 gene | **Stat5a** | signal transducer and activator of transcription 5A |
| **2210409E12Rik** | transcription elongation factor B (SIII), polypeptide 2 pseudogene | **Suz12** | suppressor of zeste 12 homolog (Drosophila) |
| **3010026O09Rik** | RIKEN cDNA 3010026O09 gene | **Tekt1** | tektin 1 |
| **4933427D14Rik** | RIKEN cDNA 4933427D14 gene | **Tex19.2** | testis expressed gene 19.2 |
| **Akap1** | A kinase (PRKA) anchor protein 1 | **Trim41** | tripartite motif-containing 41 |
| **Appbp2** | amyloid beta precursor protein (cytoplasmic tail) binding protein 2 | **Ube2b** | ubiquitin-conjugating enzyme E2B, RAD6 homology (S. cerevisiae) |
| **Aspa** | aspartoacylase | **Urgcp** | upregulator of cell proliferation |
| **Atp1b2** | ATPase, Na+/K+ transporting, beta 2 polypeptide | **Utp6** | UTP6, small subunit (SSU) processome component, homolog (yeast) |
| **C1d** | C1D nuclear receptor co-repressor | **Vps53** | vacuolar protein sorting 53 (yeast) |
| **Cbx4** | chromobox homolog 4 (Drosophila Pc class) | **Wrap53** | WD repeat containing, antisense to TP53 |
| **Cct4** | chaperonin containing Tcp1, subunit 4 (delta) | **Wsb1** | WD repeat and SOCS box-containing 1 |
| **Cct4** | chaperonin containing Tcp1, subunit 4 (delta) | **Xbp1** | X-box binding protein 1 |
| **Cpeb4** | cytoplasmic polyadenylation element binding protein 4 | **Chromosome 12** | |
| **Ddc** | dopa decarboxylase | **Frmd6** | FERM domain containing 6 |
| **Derl2** | Der1-like domain family, member 2 | **Gtf2a1** | general transcription factor II A, 1 |
| **Dynll2** | dynein light chain LC8-type 2 | **Sec23a** | SEC23A (S. cerevisiae) |
| **Fam183b** | family with sequence similarity 183, member B | **Six4** | sine oculis-related homeobox 4 homolog (Drosophila) |
| **Ggnbp2** | Gametogenetin binding protein 2 | **Uevld** | UEV and lactate/malate dehyrogenase domains |
| **H2afv** | H2A histone family, member V | **Chromosome 13** | |
| **Krt12** | keratin 12 | **Akr1c12/c13** | aldo-keto reductase family 1, member C12 /C13 |
| **Lig3** | ligase III, DNA, ATP-dependent | **Ccno** | cyclin O |
| **Llgl2** | lethal giant larvae homolog 2 (Drosophila) | **Fbp2** | fructose bisphosphatase 2 |
| **Lsmd1** | LSM domain containing 1 | **Gadd45g** | growth arrest and DNA-damage-inducible 45 gamma |
| **Med13** | mediator complex subunit 13 | **Hist1h2bp** | histone cluster 1, H2bp /// similar to Hist1h2bj protein |
| **Meis1** | Meis homeobox 1 | **Hist1h1c** | histone cluster 1, H1c |
| **Mis12** | MIS12 homolog (yeast) | **Hist1h2bc** | histone cluster 1, H2bc |
| **Mrm1** | mitochondrial rRNA methyltransferase 1 homolog (S. cerevisiae) | **Id4** | inhibitor of DNA binding 4 |
| **Mrps24** | mitochondrial ribosomal protein S24 | **Lyst** | lysosomal trafficking regulator |
| **Nbr1** | neighbor of Brca1 gene 1 | **Nid1** | nidogen 1 |
| **Ncor1** | nuclear receptor co-repressor 1 | **Nlrp4f** | NLR family, pyrin domain containing 4F |
| **Pecam1** | platelet/endothelial cell adhesion molecule 1 | **Ppap2a** | phosphatidic acid phosphatase type 2A |
| **Sept8** | septin 8 | **Chromosome 14** | |
| **Slc35b1** | solute carrier family 35, member B1 | **Gm7233** | predicted gene 7233 |
| **Slc43a2** | solute carrier family 43, member 2 | **D14Ertd436e** | DNA segment, Chr 14, ERATO Doi 436, expressed |
| **Slc46a1** | solute carrier family 46, member 1 | **Defb42** | defensin beta 42 |
| **Smtnl2** | smoothelin-like 2 | **LOC624112** | hypothetical protein LOC624112 |

## Table 4. Up-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis. (cont.)

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 14 (cont.)** | | **Chromosome 19 (cont.)** | |
| **Hesx1** | homeobox gene expressed in ES cells | **Crtac1** | cartilage acidic protein 1 |
| **Sox21** | SRY-box containing gene 21 | **Ldhb** | lactate dehydrogenase B |
| **Chromosome 15** | | **Mbl2** | mannose-binding lectin (protein C) 2 |
| **Aard** | alanine and arginine rich domain containing protein | **Mlana** | melan-A |
| **Creld2** | cysteine-rich with EGF-like domains 2 | **Ms4a10** | membrane-spanning 4-domains, subfamily A, member 10 |
| **Dab2** | disabled homolog 2 (Drosophila) | **Pga5** | pepsinogen 5, group I |
| **Gtsf1** | gametocyte specific factor 1 | **Saps3** | SAPS domain family, member 3 |
| **Myo10** | myosin X | **Chromosome X** | |
| **Sepp1** | selenoprotein P, plasma, 1 | **1700013H16Rik** | RIKEN cDNA 1700013H16 gene |
| **Chromosome 16** | | **4933402E13Rik** | RIKEN cDNA 4933402E13 gene |
| **Abcc5** | ATP-binding cassette, sub-family C (CFTR/MRP), member 5 | **Abcd1** | ATP-binding cassette, sub-family D (ALD), member 1 |
| **Alcam** | activated leukocyte cell adhesion molecule | **Pramel3** | preferentially expressed antigen in melanoma-like 3 |
| **Brwd1** | bromodomain and WD repeat domain containing 1 | **Dcaf12l1** | DDB1 and CUL4 associated factor 12-like 1 |
| **Cd200** | CD200 antigen | **Dmrtc1c** | DMRT-like family C1c /// DMRT-like family C1c2 |
| **Lrrc15** | leucine rich repeat containing 15 | **Efhc2** | EF-hand domain (C-terminal) containing 2 |
| **Chromosome 17** | | **Fgd1** | FYVE, RhoGEF and PH domain containing 1 |
| **Cdc42ep3** | CDC42 effector protein (Rho GTPase binding) 3 | **Fthl17** | ferritin, heavy polypeptide-like 17 |
| **Dazl** | deleted in azoospermia-like | **Gm7365** | predicted gene 2012 |
| **Lclat1** | lysocardiolipin acyltransferase 1 | **Luzp4 /// Ott** | leucine zipper protein 4 /// ovary testis transcribed |
| **Ly6g6e** | lymphocyte antigen 6 complex, locus G6E | **Gm9** | predicted gene 9 |
| **Qpct** | glutaminyl-peptide cyclotransferase (glutaminyl cyclase) | **Hdac6** | histone deacetylase 6 |
| **Tnfrsf21** | tumor necrosis factor receptor superfamily, member | **Itm2a** | integral membrane protein 2A |
| **Tulp4** | tubby like protein 4 | **Scml2** | sex comb on midleg-like 2 (Drosophila) |
| **Ypel5** | yippee-like 5 (Drosophila) | **Nxf3** | nuclear RNA export factor 3 |
| **Chromosome 18** | | **Rhox13** | reproductive homeobox 13 |
| **Atg12** | autophagy-related 12 (yeast) | **Rhox2a** | reproductive homeobox 2A |
| **Cidea** | cell death-inducing DNA fragmentation factor, alpha subunit-like effector A | **Rhox4** | reproductive homeobox 4A |
| **Cndp2** | CNDP dipeptidase 2 (metallopeptidase M20 family) | **Tsx** | testis specific X-linked gene |
| **Dpysl3** | dihydropyrimidinase-like 3 | **Usp9x** | ubiquitin specific peptidase 9, X chromosome |
| **Prrc1** | proline-rich coiled-coil 1 | **Xlr3a/b/c** | X-linked lymphocyte-regulated 3A/3B/3C |
| **Slc12a2** | solute carrier family 12, member 2 | **Xlr4a/b/c** | X-linked lymphocyte-regulated 4A/4B/4C |
| **Chromosome 19** | | **Zfp711** | zinc finger protein 711 |
| **Cpn1** | carboxypeptidase N, polypeptide 1 | | |

**Table 5. Down-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis.**

Genes were identified using two-way ANOVA contrasts N1 vs. WT, N1 vs. Rescue and N2 vs. WT ($p<0.0001$) and fold changes $\leq -1.5$.

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 1** | | **Chromosome 3** | |
| **Mtm1** | X-linked myotubular myopathy gene 1 | **Adar** | adenosine deaminase, RNA-specific |
| **Pstpip2** | proline-serine-threonine phosphatase-interacting protein 2 | **Bdh2** | 3-hydroxybutyrate dehydrogenase, type 2 |
| **Adam23** | a disintegrin and metallopeptidase domain 23 | **Fubp1** | far upstream element (FUSE) binding protein 1 |
| **Aldh9a1** | aldehyde dehydrogenase 9, subfamily A1 | **Isg20l2** | interferon stimulated exonuclease gene 20-like 2 |
| **Ccnt2** | cyclin T2 | **Nras** | neuroblastoma ras oncogene |
| **Chml** | choroideremia-like | **Mllt11** | myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11 |
| **Gas5** | growth arrest specific 5 | **Phf17** | PHD finger protein 17 |
| **Snord47** | small nucleolar RNA, C/D box 47 | **Pitx2** | paired-like homeodomain transcription factor 2 |
| **Gbx2** | gastrulation brain homeobox 2 | **Ppm1j** | protein phosphatase 1J |
| **Gli2** | GLI-Kruppel family member GLI2 | **Prpf3** | PRP3 pre-mRNA processing factor 3 homolog (yeast) |
| **Inhbb** | inhibin beta-B | **Psrc1** | proline/serine-rich coiled-coil 1 |
| **Slc39a10** | solute carrier family 39 (zinc transporter), member 10 | **Rabggtb** | RAB geranylgeranyl transferase, b subunit |
| **Sulf1** | sulfatase 1 | **Sars** | seryl-aminoacyl-tRNA synthetase |
| **Tuba4a** | tubulin, alpha 4A | **Snhg8** | small nucleolar RNA host gene 8 |
| **Chromosome 2** | | **Stmn2** | stathmin-like 2 |
| **1500012F01Rik** | RIKEN cDNA 1500012F01 gene | **Trim2** | tripartite motif-containing 2 |
| **Accsl** | 1-aminocyclopropane-1-carboxylate synthase homolog (Arabidopsis)(non-functional)-like | **Chromosome 4** | |
| **Cdca7** | cell division cycle associated 7 | **D4Wsu53e** | DNA segment, Chr 4, Wayne State University 53, expressed |
| **Chac1** | ChaC, cation transport regulator-like 1 (E. coli) | **Ddx58** | DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 |
| **Csrnp3** | cysteine-serine-rich nuclear protein 3 | **Lck** | lymphocyte protein tyrosine kinase |
| **Ell3** | elongation factor RNA polymerase II-like 3 | **Rcc1** | regulator of chromosome condensation 1 |
| **Glo1** | glyoxalase 1 | **Snhg12** | small nucleolar RNA host gene 12 |
| **Id1** | inhibitor of DNA binding 1 | **Snhg3** | small nucleolar RNA host gene (non-protein coding) 3 |
| **Mcm10** | minichromosome maintenance deficient 10 (S. cerevisiae) | **Tomm5** | translocase of outer mitochondrial membrane 5 homolog (yeast) |
| **Mcts2** | malignant T cell amplified sequence 2 | **Chromosome 5** | |
| **Mettl11a** | methyltransferase like 11A | **2210408K08Rik** | RIKEN cDNA 2210408K08 gene |
| **Pdk1** | pyruvate dehydrogenase kinase, isoenzyme 1 | **2700023E23Rik** | RIKEN cDNA 2700023E23 gene |
| **Pim2** | proviral integration site 2 | **Atp10d** | ATPase, class V, type 10D |
| **Plcb4** | phospholipase C, beta 4 | **Cabp1** | calcium binding protein 1 |
| **Qser1** | glutamine and serine rich 1 | **Cnpy1** | canopy 1 homolog (zebrafish) |
| **Rpl12** | ribosomal protein L12 | **Crmp1** | collapsin response mediator protein 1 |
| **Snhg7** | Small nucleolar RNA host gene (non-protein coding) 7 | **Fgfbp1** | fibroblast growth factor binding protein 1 |
| **Spred1** | sprouty protein with EVH-1 domain 1, related sequence | **Gtf2i** | general transcription factor II I |
| **Surf2** | surfeit gene 2 | **Pla2g1b** | phospholipase A2, group IB, pancreas |
| **Tdh** | L-threonine dehydrogenase | **Slc7a1** | solute carrier family 7 (cationic amino acid transporter, y+ system), member 1 |

## Table 5. Down-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis. (cont.)

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 6** | | **Chromsome 8 (cont.)** | |
| **Uba6** | ubiquitin-like modifier activating enzyme 6 | **Pgls** | 6-phosphogluconolactonase |
| **8430410A17Rik** | RIKEN cDNA 8430410A17 gene | **Prdx2** | peroxiredoxin 2 |
| **Bcat1** | branched chain aminotransferase 1, cytosolic | **Setd6** | SET domain containing 6 |
| **Bid** | BH3 interacting domain death agonist | **Tfdp1** | transcription factor Dp 1 |
| **Fgd5** | FYVE, RhoGEF and PH domain containing 5 | **Xkr5** | X Kell blood group precursor-related family, member 5 |
| **Foxp1** | forkhead box P1 | **Zfp423** | zinc finger protein 423 |
| **Gars** | glycyl-tRNA synthetase | **Chromosome 9** | |
| **Isy1** | ISY1 splicing factor homolog (S. cerevisiae) | **2610101N10Rik** | RIKEN cDNA 2610101N10 gene |
| **Mthfd2** | methylenetetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase | **2810026P18Rik** | RIKEN cDNA 2810026P18 gene |
| **Pon2** | paraoxonase 2 | **A230050P20Rik** | RIKEN cDNA A230050P20 gene |
| **Tia1** | cytotoxic granule-associated RNA binding protein 1 | **Ap1m2** | adaptor protein complex AP-1, mu 2 subunit |
| **Tpi1** | triosephosphate isomerase 1 | **AW552889** | expressed sequence AW552889 |
| **Trh** | thyrotropin releasing hormone | **Calml4** | calmodulin-like 4 |
| **Chromosome 7** | | **Exosc7** | exosome component 7 |
| **Aplp1** | amyloid beta (A4) precursor-like protein 1 | **Hspa8** | heat shock protein |
| **Igf2** | insulin-like growth factor 2 | **Ppan** | peter pan homolog (Drosophila) |
| **Ino80e** | INO80 complex subunit E | **Rps25** | ribosomal protein S25 |
| **Nupr1** | nuclear protein 1 | **Slc25a38** | solute carrier family 25, member 38 |
| **Rps9** | ribosomal protein S9 | **Taf1d** | TATA box binding protein (Tbp)-associated factor, RNA polymerase I, D |
| **Chromosome 8** | | **Tagln** | transgelin |
| **Aars** | alanyl-tRNA synthetase | **Chromosome 10** | |
| **Ankrd10** | ankyrin repeat domain 10 | **Cdk4** | cyclin-dependent kinase 4 |
| **Aprt** | adenine phosphoribosyl transferase | **Cpsf6** | cleavage and polyadenylation specific factor 6 |
| **Arl2bp** | ADP-ribosylation factor-like 2 binding protein | **Gstt2** | glutathione S-transferase, theta 2 |
| **Arrdc2** | arrestin domain containing 2 | **Pcbd1** | pterin 4 alpha carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (TCF1) 1 |
| **Cotl1** | coactosin-like 1 (Dictyostelium) | **Ric8b** | resistance to inhibitors of cholinesterase 8 homolog B (C. elegans) |
| **Dhodh** | dihydroorotate dehydrogenase | **Rpl41** | ribosomal protein L41 |
| **Gm10664** | predicted gene 10664 | **Timm8a1** | translocase of inner mitochondrial membrane 8 homolog a1 (yeast) |
| **Herpud1** | homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1 | **Chromosome 11** | |
| **Ier2** | immediate early response 2 | **Aldoc** | aldolase C, fructose-bisphosphate |
| **Ifi30** | interferon gamma inducible protein 30 | **Bahcc1** | BAH domain and coiled-coil containing 1 |
| **Slc7a5** | solute carrier family 7 (cationic amino acid transporter, y+ system), member 5 | **Coasy** | Coenzyme A synthase |
| **Lphn1** | latrophilin 1 | **Glul** | glutamate-ammonia ligase (glutamine synthetase) |
| **Naf1** | nuclear assembly factor 1 homolog (S. cerevisiae) | **Grb7** | growth factor receptor bound protein 7 |
| **Ncan** | neurocan | **Ntn1** | netrin 1 |

## Table 5. Down-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis. (cont.)

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 11 (cont.)** | | **Chromosome 15** | |
| **Psme2** | proteasome (prosome, macropain) 28 subunit, beta | **Arhgap8** | Rho GTPase activating protein 8 |
| **Slc16a3** | solute carrier family 16 (monocarboxylic acid transporters), member 3 | **Atf4** | activating transcription factor 4 |
| **Tk1** | thymidine kinase 1 | **Rpl3** | ribosomal protein L3 |
| **Chromosome 12** | | **Hnrnpa1** | heterogeneous nuclear ribonucleoprotein A1 |
| **4933433P14Rik** | RIKEN cDNA 4933433P14 gene | **Mtss1** | metastasis suppressor 1 |
| **Alkbh1** | alkB, alkylation repair homolog 1 (E. coli) | **Myg1** | melanocyte proliferating gene 1 |
| **Bnip3l** | BCL2/adenovirus E1B interacting protein 3-like | **Naprt1** | nicotinate phosphoribosyltransferase domain containing 1 |
| **Crip1** | cysteine-rich protein 1 (intestinal) | **Pvt1** | plasmacytoma variant translocation 1 |
| **G2e3** | G2/M-phase specific E3 ubiquitin ligase | | |
| **Gm71** | predicted gene 71 | **Smagp** | small cell adhesion glycoprotein |
| **Itpk1** | inositol 1,3,4-triphosphate 5/6 kinase | **Zfp740** | zinc finger protein 740 |
| **Chromosome 13** | | **Chromosome 16** | |
| **AA987161** | expressed sequence AA987161 | **2510002D24Rik** | RIKEN cDNA 2510002D24 gene |
| **AW209491** | expressed sequence AW209491 | **Bdh1** | 3-hydroxybutyrate dehydrogenase, type 1 |
| **Cdc14b** | CDC14 cell division cycle 14 homolog B (S. cerevisiae) | **Cbr3** | carbonyl reductase 3 |
| **E2f3** | E2F transcription factor 3 | **Dtx3l** | deltex 3-like (Drosophila) |
| **Emb** | embigin | **Gap43** | growth associated protein 43 |
| **F2rl1** | coagulation factor II (thrombin) receptor-like 1 | **Nit2** | nitrilase family, member 2 |
| **Idi1** | isopentenyl-diphosphate delta isomerase | **Srl** | sarcalumenin |
| **Iqgap2** | IQ motif containing GTPase activating protein 2 | **Chromosome 17** | |
| **Sfxn1** | sideroflexin 1 | **1110038B12Rik** | RIKEN cDNA 1110038B12 gene |
| **Chromosome 14** | | **1700097N02Rik** | RIKEN cDNA 1700097N02 gene |
| **Adk** | adenosine kinase | **1700097N02Rik** | RIKEN cDNA 1700097N02 gene |
| **AU020094** | expressed sequence AU020094 | **2610019E17Rik** | RIKEN cDNA 2610019E17 gene |
| **Bmp4** | bone morphogenetic protein 4 | **Bst2** | bone marrow stromal cell antigen 2 |
| **Cryl1** | crystallin, lambda 1 | **Crb3** | crumbs homolog 3 (Drosophila) |
| **Fgf17** | fibroblast growth factor 17 | **Dnahc8** | dynein, axonemal, heavy chain 8 |
| **Gng2** | guanine nucleotide binding protein (G protein), gamma 2 | **Ehmt2** | euchromatic histone lysine N-methyltransferase 2 |
| **Gnl3** | guanine nucleotide binding protein-like 3 (nucleolar) | **Gm7325** | predicted gene 7325 |
| **Pnp** | purine-nucleoside phosphorylase | **H2-D1** | histocompatibility 2, D region locus 1 |
| **Lpar6** | lysophosphatidic acid receptor 6 | **Ltb** | lymphotoxin B |
| **N6amt2** | N-6 adenine-specific DNA methyltransferase 2 (putative) | **Ppil1** | peptidylprolyl isomerase (cyclophilin)-like 1 |
| **Pck2** | phosphoenolpyruvate carboxykinase 2 (mitochondrial) | **Psmb8** | proteasome (prosome, macropain) subunit, beta type 8 (large multifunctional peptidase 7) |
| **Ppif** | peptidylprolyl isomerase F (cyclophilin F) | **Srbd1** | S1 RNA binding domain 1 |
| **Rnase4** | ribonuclease, RNase A family 4 | **Stap2** | signal transducing adaptor family member 2 |
| **Slc7a7** | solute carrier family 7 (cationic amino acid transporter, y+ system), member 7 | **T** | brachyury |
| **Tgm1** | transglutaminase 1, K polypeptide | **Tubb5** | tubulin, beta 5 |

**Table 5. Down-regulated Cfp1$^{(-/-)}$ and rescued genes submitted for functional annotation analysis. (cont.)**

| Gene Name | Description | Gene Name | Description |
|---|---|---|---|
| **Chromosome 18** | | **Chromosome 19 (cont.)** | |
| **Celf4** | CUGBP, Elav-like family member 4 | **Nolc1** | nucleolar and coiled-body phosphoprotein 1 |
| **Cxxc1** | CXXC finger 1 (PHD domain) | **Pola2** | polymerase (DNA directed), alpha 2 |
| **Impact** | imprinted and ancient | **Psat1** | phosphoserine aminotransferase 1 |
| **Nars** | asparaginyl-tRNA synthetase | **Rasgrp2** | RAS, guanyl releasing protein 2 |
| **Spry4** | sprouty homolog 4 (Drosophila) | **Scd2** | stearoyl-Coenzyme A desaturase 2 |
| **Zfp532** | zinc finger protein 532 | **Snhg1** | small nucleolar RNA host gene (non-protein coding) 1 |
| **Chromosome 19** | | **Sssca1** | Sjogren's syndrome/scleroderma autoantigen 1 homolog (human) |
| **5033414D02Rik** | RIKEN cDNA 5033414D02 gene | **Tmem180** | transmembrane protein 180 |
| **5730408K05Rik** | RIKEN cDNA 5730408K05 gene | **Chromosome X** | |
| **Aldh18a1** | aldehyde dehydrogenase 18 family, member A1 | **Mageh1** | melanoma antigen, family H, 1 |
| **Gna14** | guanine nucleotide binding protein, alpha 14 | **Mid1ip1** | Mid1 interacting protein 1 (gastrulation specific G12-like (zebrafish)) |
| **Gng3** | guanine nucleotide binding protein (G protein), gamma 3 | **Rbm3** | RNA binding motif protein 3 |
| **Ina** | internexin neuronal intermediate filament protein, alpha | **Utp14a** | UTP14, U3 small nucleolar ribonucleoprotein, homolog A (yeast) |
| **Ms4a4d** | membrane-spanning 4-domains, subfamily A, member 4D | **Xist** | inactive X specific transcripts |

**Table 6. Gene Ontology biological process associations for Cfp1$^{(-/-)}$ down-regulated genes that rescued.**

| Term for down regulated genes | Count | % | p-Value |
|---|---|---|---|
| GO:0009987~cellular process | 132 | 57.641921 | 1.25E-05 |
| GO:0008152~metabolic process | 100 | 43.668122 | 0.0021982 |
| GO:0044237~cellular metabolic process | 90 | 39.30131 | 3.51E-04 |
| GO:0044238~primary metabolic process | 87 | 37.991266 | 0.0118329 |
| GO:0006807~nitrogen compound metabolic process | 56 | 24.454148 | 6.17E-04 |
| GO:0006139~nucleobase, nucleoside, nucleotide and nucleic acid metabolic process | 49 | 21.39738 | 0.0031946 |
| GO:0044249~cellular biosynthetic process | 49 | 21.39738 | 0.0062893 |
| GO:0010467~gene expression | 42 | 18.340611 | 0.0144805 |
| GO:0019222~regulation of metabolic process | 42 | 18.340611 | 0.0958428 |
| GO:0010468~regulation of gene expression | 37 | 16.157205 | 0.0579591 |
| GO:0048518~positive regulation of biological process | 27 | 11.790393 | 0.019821 |
| GO:0048522~positive regulation of cellular process | 26 | 11.353712 | 0.0084352 |
| GO:0048519~negative regulation of biological process | 25 | 10.917031 | 0.0164717 |
| GO:0048523~negative regulation of cellular process | 22 | 9.6069869 | 0.0316062 |
| GO:0016070~RNA metabolic process | 18 | 7.860262 | 0.0015624 |
| GO:0006082~organic acid metabolic process | 17 | 7.4235808 | 1.28E-04 |
| GO:0042127~regulation of cell proliferation | 17 | 7.4235808 | 4.61E-04 |
| GO:0006915~apoptosis | 16 | 6.9868996 | 2.98E-04 |
| GO:0012501~programmed cell death | 16 | 6.9868996 | 3.57E-04 |
| GO:0019752~carboxylic acid metabolic process | 16 | 6.9868996 | 4.08E-04 |
| GO:0043436~oxoacid metabolic process | 16 | 6.9868996 | 4.08E-04 |
| GO:0042180~cellular ketone metabolic process | 16 | 6.9868996 | 5.28E-04 |
| GO:0010941~regulation of cell death | 16 | 6.9868996 | 0.002075 |
| GO:0042981~regulation of apoptosis | 15 | 6.5502183 | 0.0045818 |
| GO:0006396~RNA processing | 14 | 6.1135371 | 0.001579 |
| GO:0031325~positive regulation of cellular metabolic process | 14 | 6.1135371 | 0.0340832 |
| GO:0009893~positive regulation of metabolic process | 14 | 6.1135371 | 0.0504446 |
| GO:0010604~positive regulation of macromolecule metabolic process | 13 | 5.6768559 | 0.0603003 |
| GO:0009888~tissue development | 13 | 5.6768559 | 0.0632021 |
| GO:0009966~regulation of signal transduction | 13 | 5.6768559 | 0.07774 |
| GO:0010557~positive regulation of macromolecule biosynthetic process | 12 | 5.2401747 | 0.0416776 |
| GO:0031328~positive regulation of cellular biosynthetic process | 12 | 5.2401747 | 0.0529787 |
| GO:0009891~positive regulation of biosynthetic process | 12 | 5.2401747 | 0.05581 |
| GO:0044085~cellular component biogenesis | 12 | 5.2401747 | 0.0986113 |
| GO:0044271~nitrogen compound biosynthetic process | 11 | 4.8034934 | 0.002545 |
| GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process | 11 | 4.8034934 | 0.0689631 |
| GO:0051173~positive regulation of nitrogen compound metabolic process | 11 | 4.8034934 | 0.0808926 |
| GO:0009892~negative regulation of metabolic process | 11 | 4.8034934 | 0.0873094 |
| GO:0006519~cellular amino acid and derivative metabolic process | 10 | 4.3668122 | 0.0072935 |
| GO:0006066~alcohol metabolic process | 10 | 4.3668122 | 0.0230176 |
| GO:0045941~positive regulation of transcription | 10 | 4.3668122 | 0.0962266 |
| GO:0043066~negative regulation of apoptosis | 9 | 3.930131 | 0.0063336 |
| GO:0043069~negative regulation of programmed cell death | 9 | 3.930131 | 0.007139 |
| GO:0060548~negative regulation of cell death | 9 | 3.930131 | 0.00731 |
| GO:0044106~cellular amine metabolic process | 9 | 3.930131 | 0.009401 |
| GO:0055086~nucleobase, nucleoside and nucleotide metabolic process | 9 | 3.930131 | 0.0124141 |
| GO:0008284~positive regulation of cell proliferation | 9 | 3.930131 | 0.0167123 |
| GO:0006412~translation | 9 | 3.930131 | 0.0307495 |

**Table 6. Gene Ontology biological process associations for Cfp1$^{(-/-)}$ down-regulated genes that rescued. (cont.)**

| Term for down regulated genes | Count | % | p-Value |
|---|---|---|---|
| GO:0009308~amine metabolic process | 9 | 3.930131 | 0.0398702 |
| GO:0006520~cellular amino acid metabolic process | 8 | 3.4934498 | 0.0064871 |
| GO:0006753~nucleoside phosphate metabolic process | 8 | 3.4934498 | 0.0238081 |
| GO:0010942~positive regulation of cell death | 8 | 3.4934498 | 0.0261892 |
| GO:0006397~mRNA processing | 8 | 3.4934498 | 0.0314299 |
| GO:0032787~monocarboxylic acid metabolic process | 8 | 3.4934498 | 0.0348911 |
| GO:0048534~hemopoietic or lymphoid organ development | 8 | 3.4934498 | 0.0432378 |
| GO:0046483~heterocycle metabolic process | 8 | 3.4934498 | 0.047492 |
| GO:0002520~immune system development | 8 | 3.4934498 | 0.0535684 |
| GO:0016071~mRNA metabolic process | 8 | 3.4934498 | 0.0592693 |
| GO:0044262~cellular carbohydrate metabolic process | 8 | 3.4934498 | 0.082595 |
| GO:0006006~glucose metabolic process | 7 | 3.0567686 | 0.0055005 |
| GO:0019318~hexose metabolic process | 7 | 3.0567686 | 0.0132642 |
| GO:0005996~monosaccharide metabolic process | 7 | 3.0567686 | 0.0227693 |
| GO:0034660~ncRNA metabolic process | 7 | 3.0567686 | 0.028834 |
| GO:0008283~cell proliferation | 7 | 3.0567686 | 0.0650799 |
| GO:0043065~positive regulation of apoptosis | 7 | 3.0567686 | 0.0660918 |
| GO:0043068~positive regulation of programmed cell death | 7 | 3.0567686 | 0.0681429 |
| GO:0035295~tube development | 7 | 3.0567686 | 0.0835298 |
| GO:0046394~carboxylic acid biosynthetic process | 6 | 2.6200873 | 0.0231658 |
| GO:0006732~coenzyme metabolic process | 6 | 2.6200873 | 0.024439 |
| GO:0010608~posttranscriptional regulation of gene expression | 6 | 2.6200873 | 0.0278077 |
| GO:0034654~nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process | 6 | 2.6200873 | 0.0552934 |
| GO:0051186~cofactor metabolic process | 6 | 2.6200873 | 0.0584643 |
| GO:0009119~ribonucleoside metabolic process | 5 | 2.1834061 | 0.0010015 |
| GO:0009116~nucleoside metabolic process | 5 | 2.1834061 | 0.0038468 |
| GO:0030324~lung development | 5 | 2.1834061 | 0.0387928 |
| GO:0048872~homeostasis of number of cells | 5 | 2.1834061 | 0.0409918 |
| GO:0060541~respiratory system development | 5 | 2.1834061 | 0.0543278 |
| GO:0022613~ribonucleoprotein complex biogenesis | 5 | 2.1834061 | 0.0727601 |
| GO:0046128~purine ribonucleoside metabolic process | 4 | 1.7467249 | 0.0025079 |
| GO:0008652~cellular amino acid biosynthetic process | 4 | 1.7467249 | 0.0138896 |
| GO:0043039~tRNA aminoacylation | 4 | 1.7467249 | 0.0156582 |
| GO:0045619~regulation of lymphocyte differentiation | 4 | 1.7467249 | 0.0288267 |
| GO:0009309~amine biosynthetic process | 4 | 1.7467249 | 0.0513906 |
| GO:0043523~regulation of neuron apoptosis | 4 | 1.7467249 | 0.0641299 |
| GO:0017038~protein import | 4 | 1.7467249 | 0.0680135 |
| GO:0021915~neural tube development | 4 | 1.7467249 | 0.0699946 |
| GO:0019882~antigen processing and presentation | 4 | 1.7467249 | 0.0781762 |
| GO:0006916~anti-apoptosis | 4 | 1.7467249 | 0.0802844 |
| GO:0048754~branching morphogenesis of a tube | 4 | 1.7467249 | 0.0911891 |
| GO:0006094~gluconeogenesis | 3 | 1.3100437 | 0.0216753 |
| GO:0021532~neural tube patterning | 3 | 1.3100437 | 0.0305627 |
| GO:0006090~pyruvate metabolic process | 3 | 1.3100437 | 0.0329746 |
| GO:0019319~hexose biosynthetic process | 3 | 1.3100437 | 0.0329746 |
| GO:0019362~pyridine nucleotide metabolic process | 3 | 1.3100437 | 0.0406344 |
| GO:0021983~pituitary gland development | 3 | 1.3100437 | 0.0433227 |
| GO:0046364~monosaccharide biosynthetic process | 3 | 1.3100437 | 0.0488909 |
| GO:0046165~alcohol biosynthetic process | 3 | 1.3100437 | 0.0670005 |

**Table 7. Gene Ontology biological process associations with Cfp1$^{(-/-)}$ up-regulated genes that rescued.**

| Term for up-regulated genes | Count | % | p-Value |
|---|---|---|---|
| GO:0009987~cellular process | 171 | 51.975684 | 8.27E-05 |
| GO:0008152~metabolic process | 123 | 37.386018 | 0.0463928 |
| GO:0065007~biological regulation | 120 | 36.474164 | 0.0641973 |
| GO:0044238~primary metabolic process | 110 | 33.43465 | 0.0513555 |
| GO:0044237~cellular metabolic process | 106 | 32.218845 | 0.0323534 |
| GO:0019222~regulation of metabolic process | 60 | 18.237082 | 0.0183699 |
| GO:0080090~regulation of primary metabolic process | 54 | 16.413374 | 0.0311794 |
| GO:0060255~regulation of macromolecule metabolic process | 54 | 16.413374 | 0.0344414 |
| GO:0051179~localization | 54 | 16.413374 | 0.0287142 |
| GO:0031323~regulation of cellular metabolic process | 53 | 16.109422 | 0.0828597 |
| GO:0007275~multicellular organismal development | 49 | 14.893617 | 0.0725284 |
| GO:0051234~establishment of localization | 48 | 14.589666 | 0.03942 |
| GO:0010468~regulation of gene expression | 48 | 14.589666 | 0.060479 |
| GO:0006810~transport | 48 | 14.589666 | 0.0355756 |
| GO:0010556~regulation of macromolecule biosynthetic process | 46 | 13.981763 | 0.089138 |
| GO:0048856~anatomical structure development | 44 | 13.37386 | 0.0524091 |
| GO:0045449~regulation of transcription | 44 | 13.37386 | 0.0735059 |
| GO:0048518~positive regulation of biological process | 42 | 12.765957 | 2.05E-04 |
| GO:0044267~cellular protein metabolic process | 42 | 12.765957 | 0.0648552 |
| GO:0016043~cellular component organization | 42 | 12.765957 | 0.0161893 |
| GO:0048522~positive regulation of cellular process | 37 | 11.246201 | 6.04E-04 |
| GO:0048513~organ development | 37 | 11.246201 | 0.0245824 |
| GO:0048519~negative regulation of biological process | 32 | 9.7264438 | 0.0139719 |
| GO:0051252~regulation of RNA metabolic process | 31 | 9.4224924 | 0.082593 |
| GO:0043412~biopolymer modification | 31 | 9.4224924 | 0.0220775 |
| GO:0006996~organelle organization | 31 | 9.4224924 | 0.0013132 |
| GO:0006464~protein modification process | 31 | 9.4224924 | 0.0122951 |
| GO:0006355~regulation of transcription, DNA-dependent | 31 | 9.4224924 | 0.0706858 |
| GO:0006950~response to stress | 29 | 8.8145897 | 0.0167599 |
| GO:0043687~post-translational protein modification | 28 | 8.5106383 | 0.0071279 |
| GO:0048523~negative regulation of cellular process | 27 | 8.2066869 | 0.049076 |
| GO:0009653~anatomical structure morphogenesis | 26 | 7.9027356 | 0.0401944 |
| GO:0006508~proteolysis | 24 | 7.2948328 | 0.0520837 |
| GO:0006796~phosphate metabolic process | 23 | 6.9908815 | 0.0158859 |
| GO:0006629~lipid metabolic process | 21 | 6.3829787 | 0.006044 |
| GO:0009605~response to external stimulus | 20 | 6.0790274 | 0.0052709 |
| GO:0007399~nervous system development | 20 | 6.0790274 | 0.0732855 |
| GO:0044255~cellular lipid metabolic process | 19 | 5.775076 | 5.41E-04 |
| GO:0016310~phosphorylation | 19 | 5.775076 | 0.03122 |
| GO:0010646~regulation of cell communication | 19 | 5.775076 | 0.0556288 |
| GO:0000003~reproduction | 19 | 5.775076 | 0.0087016 |

**Table 7. Gene Ontology biological process associations with Cfp1$^{(-/-)}$ up-regulated genes that rescued. (cont.)**

| Term for up-regulated genes | Count | % | p-Value |
|---|---|---|---|
| GO:0044085~cellular component biogenesis | 18 | 5.4711246 | 0.0163518 |
| GO:0031325~positive regulation of cellular metabolic process | 18 | 5.4711246 | 0.0233502 |
| GO:0009966~regulation of signal transduction | 18 | 5.4711246 | 0.0290292 |
| GO:0009893~positive regulation of metabolic process | 18 | 5.4711246 | 0.0378409 |
| GO:0051276~chromosome organization | 17 | 5.1671733 | 5.92E-04 |
| GO:0050793~regulation of developmental process | 17 | 5.1671733 | 0.0159741 |
| GO:0010604~positive regulation of macromolecule metabolic process | 17 | 5.1671733 | 0.0384304 |
| GO:0009888~tissue development | 17 | 5.1671733 | 0.0408015 |
| GO:0042127~regulation of cell proliferation | 16 | 4.8632219 | 0.0213624 |
| GO:0022607~cellular component assembly | 16 | 4.8632219 | 0.0155275 |
| GO:0009887~organ morphogenesis | 16 | 4.8632219 | 0.0445547 |
| GO:0006468~protein amino acid phosphorylation | 16 | 4.8632219 | 0.0744086 |
| GO:0006357~regulation of transcription from RNA polymerase II promoter | 16 | 4.8632219 | 0.0574909 |
| GO:0051173~positive regulation of nitrogen compound metabolic process | 15 | 4.5592705 | 0.0356077 |
| GO:0031328~positive regulation of cellular biosynthetic process | 15 | 4.5592705 | 0.0493491 |
| GO:0009891~positive regulation of biosynthetic process | 15 | 4.5592705 | 0.0528695 |
| GO:0009611~response to wounding | 15 | 4.5592705 | 0.0010832 |
| GO:0045935~positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process | 14 | 4.2553191 | 0.0555018 |
| GO:0010557~positive regulation of macromolecule biosynthetic process | 14 | 4.2553191 | 0.0705006 |
| GO:0048598~embryonic morphogenesis | 13 | 3.9513678 | 0.0105262 |
| GO:0045941~positive regulation of transcription | 13 | 3.9513678 | 0.0675957 |
| GO:0043436~oxoacid metabolic process | 13 | 3.9513678 | 0.0708097 |
| GO:0042180~cellular ketone metabolic process | 13 | 3.9513678 | 0.081852 |
| GO:0019752~carboxylic acid metabolic process | 13 | 3.9513678 | 0.0708097 |
| GO:0010628~positive regulation of gene expression | 13 | 3.9513678 | 0.0792678 |
| GO:0007420~brain development | 13 | 3.9513678 | 0.0018568 |
| GO:0007417~central nervous system development | 13 | 3.9513678 | 0.0128639 |
| GO:0006325~chromatin organization | 13 | 3.9513678 | 0.0038513 |
| GO:0006082~organic acid metabolic process | 13 | 3.9513678 | 0.0714219 |
| GO:0051246~regulation of protein metabolic process | 12 | 3.6474164 | 0.0215394 |
| GO:0006952~defense response | 12 | 3.6474164 | 0.0908784 |
| GO:0006066~alcohol metabolic process | 12 | 3.6474164 | 0.0256331 |
| GO:0003006~reproductive developmental process | 12 | 3.6474164 | 0.0028461 |
| GO:0051172~negative regulation of nitrogen compound metabolic process | 11 | 3.343465 | 0.096226 |
| GO:0048646~anatomical structure formation involved in morphogenesis | 11 | 3.343465 | 0.0592581 |
| GO:0045934~negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process | 11 | 3.343465 | 0.0906665 |
| GO:0045595~regulation of cell differentiation | 11 | 3.343465 | 0.0791025 |
| GO:0032787~monocarboxylic acid metabolic process | 11 | 3.343465 | 0.0092512 |
| GO:0019953~sexual reproduction | 11 | 3.343465 | 0.0791025 |
| GO:0016481~negative regulation of transcription | 11 | 3.343465 | 0.0646483 |
| GO:0060429~epithelium development | 10 | 3.0395137 | 0.0256687 |
| GO:0051253~negative regulation of RNA metabolic process | 10 | 3.0395137 | 0.0535652 |
| GO:0051128~regulation of cellular component organization | 10 | 3.0395137 | 0.0723915 |

**Table 7. Gene Ontology biological process associations with Cfp1$^{(-/-)}$ up-regulated genes that rescued. (cont.)**

| Term for up-regulated genes | Count | % | p-Value |
|---|---|---|---|
| GO:0048610~reproductive cellular process | 10 | 3.0395137 | 0.0015438 |
| GO:0045892~negative regulation of transcription, DNA-dependent | 10 | 3.0395137 | 0.0513362 |
| GO:0008284~positive regulation of cell proliferation | 10 | 3.0395137 | 0.0337586 |
| GO:0007167~enzyme linked receptor protein signaling pathway | 10 | 3.0395137 | 0.0271792 |
| GO:0002682~regulation of immune system process | 10 | 3.0395137 | 0.0646006 |
| GO:0060548~negative regulation of cell death | 9 | 2.7355623 | 0.0377779 |
| GO:0043069~negative regulation of programmed cell death | 9 | 2.7355623 | 0.0370824 |
| GO:0043066~negative regulation of apoptosis | 9 | 2.7355623 | 0.0334225 |
| GO:0040008~regulation of growth | 9 | 2.7355623 | 0.0470055 |
| GO:0019637~organophosphate metabolic process | 9 | 2.7355623 | 0.0062845 |
| GO:0016568~chromatin modification | 9 | 2.7355623 | 0.0313595 |
| GO:0006954~inflammatory response | 9 | 2.7355623 | 0.0243772 |
| GO:0006644~phospholipid metabolic process | 9 | 2.7355623 | 0.0039805 |
| GO:0002684~positive regulation of immune system process | 9 | 2.7355623 | 0.015327 |
| GO:0046486~glycerolipid metabolic process | 8 | 2.4316109 | 0.0040188 |
| GO:0022603~regulation of anatomical structure morphogenesis | 8 | 2.4316109 | 0.027008 |
| GO:0000122~negative regulation of transcription from RNA polymerase II promoter | 8 | 2.4316109 | 0.0693751 |
| GO:0048732~gland development | 7 | 2.1276596 | 0.0873439 |
| GO:0048584~positive regulation of response to stimulus | 7 | 2.1276596 | 0.0705596 |
| GO:0042060~wound healing | 7 | 2.1276596 | 0.0080451 |
| GO:0010627~regulation of protein kinase cascade | 7 | 2.1276596 | 0.0341475 |
| GO:0002009~morphogenesis of an epithelium | 7 | 2.1276596 | 0.053319 |
| GO:0001655~urogenital system development | 7 | 2.1276596 | 0.0265355 |
| GO:0051604~protein maturation | 6 | 1.8237082 | 0.0169203 |
| GO:0051347~positive regulation of transferase activity | 6 | 1.8237082 | 0.0595808 |
| GO:0050778~positive regulation of immune response | 6 | 1.8237082 | 0.0611103 |
| GO:0045087~innate immune response | 6 | 1.8237082 | 0.0257038 |
| GO:0043408~regulation of MAPKKK cascade | 6 | 1.8237082 | 0.0149292 |
| GO:0042110~T cell activation | 6 | 1.8237082 | 0.0347324 |
| GO:0016485~protein processing | 6 | 1.8237082 | 0.0131015 |
| GO:0007281~germ cell development | 6 | 1.8237082 | 0.0206169 |
| GO:0002526~acute inflammatory response | 6 | 1.8237082 | 0.0085385 |
| GO:0002460~adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains | 6 | 1.8237082 | 0.0099126 |
| GO:0002250~adaptive immune response | 6 | 1.8237082 | 0.0099126 |
| GO:0070647~protein modification by small protein conjugation or removal | 5 | 1.5197568 | 0.0788601 |
| GO:0065004~protein-DNA complex assembly | 5 | 1.5197568 | 0.0305619 |
| GO:0051605~protein maturation by peptide bond cleavage | 5 | 1.5197568 | 0.018378 |
| GO:0050878~regulation of body fluid levels | 5 | 1.5197568 | 0.0498981 |
| GO:0050817~coagulation | 5 | 1.5197568 | 0.0234518 |
| GO:0045017~glycerolipid biosynthetic process | 5 | 1.5197568 | 0.0125202 |
| GO:0042493~response to drug | 5 | 1.5197568 | 0.0482186 |
| GO:0030879~mammary gland development | 5 | 1.5197568 | 0.0449553 |

**Table 7. Gene Ontology biological process associations with Cfp1$^{(-/-)}$ up-regulated genes that rescued. (cont.)**

| Term for up-regulated genes | Count | % | p-Value |
|---|---|---|---|
| GO:0030217~T cell differentiation | 5 | 1.5197568 | 0.0305619 |
| GO:0008654~phospholipid biosynthetic process | 5 | 1.5197568 | 0.0403008 |
| GO:0007599~hemostasis | 5 | 1.5197568 | 0.0245586 |
| GO:0007596~blood coagulation | 5 | 1.5197568 | 0.0234518 |
| GO:0006650~glycerophospholipid metabolic process | 5 | 1.5197568 | 0.0482186 |
| GO:0006333~chromatin assembly or disassembly | 5 | 1.5197568 | 0.0900312 |
| GO:0002449~lymphocyte mediated immunity | 5 | 1.5197568 | 0.0305619 |
| GO:0002443~leukocyte mediated immunity | 5 | 1.5197568 | 0.0498981 |
| GO:0002253~activation of immune response | 5 | 1.5197568 | 0.0449553 |
| GO:0046777~protein amino acid autophosphorylation | 4 | 1.2158055 | 0.0832287 |
| GO:0046164~alcohol catabolic process | 4 | 1.2158055 | 0.0803244 |
| GO:0043410~positive regulation of MAPKKK cascade | 4 | 1.2158055 | 0.0364083 |
| GO:0043406~positive regulation of MAP kinase activity | 4 | 1.2158055 | 0.0638737 |
| GO:0042472~inner ear morphogenesis | 4 | 1.2158055 | 0.0861778 |
| GO:0030850~prostate gland development | 4 | 1.2158055 | 0.0224468 |
| GO:0016567~protein ubiquitination | 4 | 1.2158055 | 0.0832287 |
| GO:0009880~embryonic pattern specification | 4 | 1.2158055 | 0.016817 |
| GO:0008643~carbohydrate transport | 4 | 1.2158055 | 0.066496 |
| GO:0007179~transforming growth factor beta receptor signaling pathway | 4 | 1.2158055 | 0.0384039 |
| GO:0006959~humoral immune response | 4 | 1.2158055 | 0.0515132 |
| GO:0060742~epithelial cell differentiation involved in prostate gland development | 3 | 0.9118541 | 0.0167571 |
| GO:0060740~prostate gland epithelium morphogenesis | 3 | 0.9118541 | 0.0612312 |
| GO:0060512~prostate gland morphogenesis | 3 | 0.9118541 | 0.0654715 |
| GO:0046632~alpha-beta T cell differentiation | 3 | 0.9118541 | 0.0345672 |
| GO:0046631~alpha-beta T cell activation | 3 | 0.9118541 | 0.0491225 |
| GO:0021575~hindbrain morphogenesis | 3 | 0.9118541 | 0.0612312 |
| GO:0017015~regulation of transforming growth factor beta receptor signaling pathway | 3 | 0.9118541 | 0.0787472 |
| GO:0006958~complement activation, classical pathway | 3 | 0.9118541 | 0.0787472 |
| GO:0006898~receptor-mediated endocytosis | 3 | 0.9118541 | 0.0416027 |

**Table 8. PANTHER molecular function identified for Cfp1$^{(-/-)}$ up- and down-regulated genes that rescued.**

Cfp1$^{(-/-)}$ mis-regulated genes that rescued, significant at $p<0.0001$ and +/- 1.5 fold change were submitted for analysis as described in materials and methods.

| Terms for down-regulated genes | Gene count | % | p-Value |
|---|---|---|---|
| MF00042:Nucleic acid binding | 47 | 14.29 | 0.059401186 |
| MF00036:Transcription factor | 39 | 11.85 | 0.018856154 |
| MF00082:Transporter | 16 | 4.86 | 0.043200469 |
| MF00038:Homeobox transcription factor | 9 | 2.74 | 0.021659425 |
| MF00275:Transcription cofactor | 6 | 1.82 | 0.083071924 |
| MF00102:Protease inhibitor | 6 | 1.82 | 0.095065112 |
| MF00250:Serine protease inhibitor | 5 | 1.52 | 0.081766783 |
| MF00013:Tyrosine protein kinase receptor | 4 | 1.22 | 0.092260048 |
| MF00088:Apolipoprotein | 3 | 0.91 | 0.051576208 |
| | | | |
| **Terms for up-regulated genes** | | | |
| MF00042:Nucleic acid binding | 38 | 16.59 | 0.019296911 |
| MF00141:Hydrolase | 14 | 6.11 | 0.027990075 |
| MF00123:Oxidoreductase | 12 | 5.24 | 0.063821958 |
| MF00053:Other RNA-binding protein | 8 | 3.49 | 0.005414587 |
| MF00166:Isomerase | 7 | 3.06 | 0.003302667 |
| MF00118:Synthase and synthetase | 7 | 3.06 | 0.023620804 |
| MF00126:Dehydrogenase | 7 | 3.06 | 0.023620804 |
| MF00120:Synthetase | 6 | 2.62 | 0.002706699 |
| MF00156:Other hydrolase | 6 | 2.62 | 0.042330298 |
| MF00157:Lyase | 5 | 2.18 | 0.075028703 |
| MF00139:Acyltransferase | 4 | 1.75 | 0.0877264 |
| MF00138:Transaminase | 3 | 1.31 | 0.024511184 |
| MF00288:Amino acid transporter | 3 | 1.31 | 0.044197149 |
| MF00121:Aminoacyl-tRNA synthetase | 3 | 1.31 | 0.054309725 |
| MF00167:Epimerase/racemase | 3 | 1.31 | 0.088710546 |
| MF00098:Large G-protein | 3 | 1.31 | 0.094945307 |

**Table 9. Kegg pathways identified in Cfp1$^{(-/-)}$ up- and down-regulated genes that rescued.**

Cfp1$^{(-/-)}$ mis-regulated genes that rescued, significant at p<0.0001 and +/- 1.5 fold change were submitted for analysis as described in materials and methods.

| Down-regulated genes | Gene count | % | p-Value |
|---|---|---|---|
| mmu00970:Aminoacyl-tRNA biosynthesis | 4 | 1.746724891 | 0.025263006 |
| mmu03040:Spliceosome | 6 | 2.620087336 | 0.039386521 |
| mmu04350:TGF-beta signaling pathway | 5 | 2.183406114 | 0.042410534 |
| mmu03010:Ribosome | 5 | 2.183406114 | 0.045486254 |
| mmu00010:Glycolysis / Gluconeogenesis | 4 | 1.746724891 | 0.083572779 |
| mmu05218:Melanoma | 4 | 1.746724891 | 0.09234171 |
| **Up-regulated genes** | | | |
| mmu05010:Alzheimer's disease | 8 | 2.431611 | 0.011402 |
| mmu00565:Ether lipid metabolism | 4 | 1.215805 | 0.011572 |
| mmu00010:Glycolysis / Gluconeogenesis | 5 | 1.519757 | 0.013339 |
| mmu04610:Complement and coagulation cascades | 5 | 1.519757 | 0.018544 |
| mmu00561:Glycerolipid metabolism | 4 | 1.215805 | 0.025535 |
| mmu00564:Glycerophospholipid metabolism | 4 | 1.215805 | 0.062319 |
| mmu00350:Tyrosine metabolism | 3 | 0.911854 | 0.09795 |

**Table 10. Identification of Cfp1$^{(-/-)}$ up-regulated genes not rescued.**
Gene lists were created using p<0.05 and + 1.5 fold change cutoff for consistent up-regulation in N1 and N2 as described in materials and methods.

| Gene Symbol | Gene Description | Chr | Gene Ontology Biological Process |
|---|---|---|---|
| 3110045C21 Rik | RIKEN cDNA 3110045C21 gene | chr1 | --- |
| 4833420G17 Rik | RIKEN cDNA 4833420G17 gene | chr13 | --- |
| 6430706D22 Rik /// Hjurp | RIKEN cDNA 6430706D22 gene /// Holliday junction recognition protein | chr1 | cell cycle // chromosome segregation // centromeric core chromatin assembly |
| 6720457D02 Rik | RIKEN cDNA 6720457D02 gene | chr10 | --- |
| B230117O15 Rik | RIKEN cDNA B230117O15 gene | chr4 | --- |
| BC003267 | cDNA sequence BC003267 | chr8 | regulation of transcription, DNA-dependent |
| BC005512 /// F630007L15 Rik /// Gm6958 | cDNA sequence BC005512 /// RIKEN cDNA F630007L15 gene /// predicted gene 6958 | --- | viral infectious cycle |
| Cdv3 | carnitine deficiency-associated gene expressed in ventricle 3 | chr9 | --- |
| Clta | clathrin, light polypeptide (Lca) | chr4 | endocytosis // vesicle-mediated transport |
| Csnk1d | casein kinase 1, delta | chr11 | protein amino acid phosphorylation // Wnt receptor signaling pathway // positive regulation of proteasomal ubiquitin-dependent protein catabolic process // circadian regulation of gene expression // |
| Cyr61 | cysteine rich protein 61 | chr3 | regulation of cell growth // intussusceptive angiogenesis // chemotaxis // cell adhesion // labyrinthine layer blood vessel development |
| Dnaja2 | DnaJ (Hsp40) homolog, subfamily A, member 2 | chr8 | protein folding // response to heat // |
| Duxbl /// Gm10394 | double homeobox B-like /// predicted gene 10394 | chr14 | regulation of transcription, DNA-dependent |
| Dzip1 | DAZ interacting protein 1 | chr14 | multicellular organismal development // spermatogenesis // cell differentiation |
| Phgdh | 3-phosphoglycerate dehydrogenase | chr14 | glutamine metabolic process //multiple metabolic processes and neural development |
| Gm5124 | predicted pseudogene 5124 | chrX | --- |
| Hmgcr | 3-hydroxy-3-methylglutaryl-Coenzyme A reductase | chr13 | steroid biosynthetic process // cholesterol biosynthetic process // visual learning // embryonic development // coenzyme A metabolic process // negative regulation of MAP kinase activity |
| Hyal2 | hyaluronoglucosaminidase 2 | chr9 | carbohydrate metabolic process // hyaluronan catabolic process |

**Table 10. Identification of Cfp1$^{(-/-)}$ up-regulated genes not rescued. (cont.)**

| Gene Symbol | Gene Description | Chr | Gene Ontology Biological Process |
|---|---|---|---|
| Iqcc | IQ motif containing C | chr4 | --- |
| Khdrbs1 | KH domain containing, RNA binding, signal transduction associated 1 | chr4 | transcription // cell cycle // cell surface receptor linked signaling pathway // negative regulation of transcription // regulation of RNA export from nucleus |
| Lamc2 | laminin, gamma 2 | chr1 | cell adhesion |
| Lnx1 | ligand of numb-protein X 1 | chr5 | ubiquitin-dependent protein catabolic process // inferred from direct assay |
| Mest | mesoderm specific transcript | chr6 | regulation of lipid storage // response to retinoic acid // |
| Mypn | myopalladin | chr10 | --- |
| Nrp2 | neuropilin 2 | chr1 | neural crest cell migration // cell adhesion // multicellular organismal development // nervous system development // axon guidance // heart development // cell differentiation // |
| Ostc | oligosaccharyltransferase complex subunit | chr3 | --- |
| Parp3 | poly (ADP-ribose) polymerase family, member 3 | chr9 | protein amino acid ADP-ribosylation |
| Peg3 | paternally expressed 3 | chr7 | apoptosis |
| Pgap2 | post-GPI attachment to proteins 2 | chr7 | GPI anchor biosynthetic process // anti-apoptosis // response to DNA damage stimulus |
| Pigo | phosphatidylinositol glycan anchor biosynthesis, class O | chr4 | GPI anchor biosynthetic process // metabolic process // inferred from electronic annotation |
| Pla2g4a | phospholipase A2, group IVA (cytosolic, calcium-dependent) | chr1 | ovulation from ovarian follicle // luteolysis // positive regulation of cell proliferation // response to heat // response to hormone stimulus / |
| Plcxd2 | phosphatidylinositol-specific phospholipase C, X domain containing 2 | chr16 | lipid metabolic process |
| Prdx2 | peroxiredoxin 2 | chr8 | activation of MAPK activity // inferred from mutant phenotype // respiratory burst involved in inflammatory response // response to oxidative stress // negative regulation of T cell differentiation // thymus development // homeostasis of number of cells |
| Ptprf | protein tyrosine phosphatase, receptor type, F | chr4 | negative regulation of cytokine-mediated signaling pathway // nervous system development // negative regulation of epidermal growth factor receptor signaling pathway // positive regulation of apoptosis // multiple receptor pathways |
| Ptprg | protein tyrosine phosphatase, receptor type, G | chr14 | protein amino acid dephosphorylation // negative regulation of neuron projection development |
| Rbp1 | retinol binding protein 1, cellular | chr9 | regulation of granulocyte differentiation // inferred from mutant phenotype // response to vitamin A // retinol metabolic process |
| Reep3 | receptor accessory protein 3 | chr10 | --- |
| Riok2 | RIO kinase 2 (yeast) | chr17 | protein amino acid phosphorylation |
| Sar1a | SAR1 gene homolog A (S. cerevisiae) | chr10 | transport // intracellular protein transport // multicellular organismal development |
| Slc1a5 | solute carrier family 1 (neutral amino acid transporter), member 5 | chr7 | transport //dicarboxylic acid transport // extracellular amino acid transport // L-serine transport |

**Table 11. Identification of Cfp1$^{(-/-)}$ down-regulated genes not rescued.**
Gene lists were created using p<.05 and ≤-1.5 fold change cutoff for consistent down-regulation in N1 and N2 as described in materials and methods.

| Gene Symbol | Gene Description | Chr | Gene Ontology Biological Process |
|---|---|---|---|
| Ramp3 | receptor (calcitonin) activity modifying protein 3 | chr11 | intracellular protein transport //G-protein coupled receptor protein signaling pathway |
| 1700029I01Rik | RIKEN cDNA 1700029I01 gene | chr4 | regulation of transcription, DNA-dependent |
| 6820431F20Rik | cadherin 11 pseudogene | chr8 | --- |
| C330019L16Rik /// Gm2381 | RIKEN cDNA C330019L16 gene /// predicted gene 2381 | chr7 | cellular response to heat |
| Ccnb1ip1 | cyclin B1 interacting protein 1 | chr14 | blastocyst formation // apoptosis //meiotic metaphase I // spermatid development // organ growth //chiasma assembly |
| Ciapin1 | cytokine induced apoptosis inhibitor 1 | chr8 | apoptosis // anti-apoptosis // hemopoiesis |
| Dvl1 | dishevelled, dsh homolog 1 (Drosophila) | chr4 | regulation of neurotransmitter levels //negative regulation of protein kinase activity // multicellular organismal development // Wnt receptor signaling pathway // |
| Frg1 | FSHD region gene 1 | chr8 | rRNA processing// RNA splicing //ribosome biogenesis |
| Gprc5b | G protein-coupled receptor, family C, group 5, member B | chr7 | signal transduction // G-protein coupled receptor protein signaling pathway |
| Jam2 | junction adhesion molecule 2 | chr16 | --- |
| Nek3 | NIMA (never in mitosis gene a)-related expressed kinase 3 | chr8 | protein amino acid phosphorylation //cell cycle // mitosis // |
| Rbm45 | RNA binding motif protein 45 | chr2 | multicellular organismal development // nervous system development // cell differentiation |
| Rwdd3 | RWD domain containing 3 | chr3 | --- |
| Sec23ip | Sec23 interacting protein | chr7 | --- |
| Ska1 | spindle and kinetochore associated complex subunit 1 | chr18 | mitotic anaphase // chromosome segregation // regulation of microtubule polymerization or depolymerization |
| Slc47a1 | solute carrier family 47, member 1 | chr11 | transport // drug transmembrane transport |
| Vegfc | vascular endothelial growth factor C | chr8 | angiogenesis //positive regulation of neuroblast proliferation // positive regulation of blood vessel endothelial cell migration // positive regulation of epithelial cell proliferation |

# CHAPTER 4: DISCUSSION

We have shown that Wdr82 is essential for development, Setd1A and Setd1B both play a critical role in cellular division, and Cfp1 plays a critical role in transcription regulation through both activation and repression of genes. In addition, Cfp1 was identified as a critical regulator of genes for basic cellular processes including programmed cell death, development and multiple biosynthetic and metabolic pathways.

## 4.1 Role of Wdr82 in Development

We have demonstrated that Wdr82 is essential for embryonic development, although the earliest time point Wdr82 is required has yet to be determined. We have not yet identified homozygous null Wdr82 embryos or ES cells (data not shown) and this suggests Wdr82 may be essential for mammalian cellular viability, consistent with Swd2 requirement for yeast viability (Cheng et al., 2004). Wdr82 depletion by siRNA has been used to evaluate the role of Wdr82 in several differentiated cell lines in previous studies, yet some protein is still detectable by western blot (Wu et al., 2008). However, unicellular gametes containing a single copy of the disrupted Wdr82 allele are viable. No phenotypic differences have been observed in heterozygous mice other than the interesting over-representation of Wdr82 heterozygous offspring following Wdr82 het cross-breeding and this merits further investigation.

There is currently not enough information available to justify a solid hypothesis for the mechanism at work, but the distribution frequency of heterozygous embryos to WT suggests altered functionality in either egg or sperm as the skewed frequency

compared to an expected Mendelian frequency would be less dramatic if both were equally affected. Further investigation is needed to identify the frequency and functional impact of the disrupted Wdr82 gene in oocytes and spermatazoa followed by an evaluation of events following fertilization. Temporal differences exist in male and female gamete meiosis and X-inactivation pre- and post-fertilization that may play a role in this unusual distribution favoring heterozygotes.

We have demonstrated functional β-galactosidase by theWdr82 fusion transcript through X-gal staining, yet it is unknown if the truncated Wdr82 fragment has any functionality. Previous studies with Swd2 truncations suggest it is unlikely as no viable yeast could be identified with any truncation of the Swd2 protein (Cheng et al., 2004; Roguev et al., 2004; Dichtl et al., 2004). However, partial Wdr82 may still be able to interact with one or more of the multiple proteins it has been found to associate with. This could be determined through immuno-precipitation of the β-galactosidase fusion protein, followed by isolation on page gel and identification of binding partners with mass spectrometry. The functional β-galactosidase will enable further studies to identify the developmental stages where Wdr82 is expressed, including identification of gamete distribution and adult tissue assessment by staining with X-gal and/or antibody.

Wdr82 het ES cells exhibit reduced amounts of Wdr82 protein compared to WT, indicating WT levels are not required for survival and heterozygous mice appear normal. The difference observed in differentiation dynamics suggests that altered Wdr82 levels does have a functional impact and may play a role in cellular division and differentiation, yet this observation may be due to clonal variation. The impact of Wdr82 protein levels can be verified by increasing Wdr82 levels in the Wdr82 het cell line with transfection of

Wdr82 cDNA, and/or reduced Wdr82 levels in WT ES cells by siRNA and observation of differential dynamics under these altered conditions.

### 4.2 Role of Setd1B and Setd1A in cellular division

Over-expression of either full length Setd1 or truncated Setd1B resulted in fewer viable cells after 4 days growth than vector, demonstrating a significant role for both Setd1 proteins in cellular division. With and without the ability to bind Wdr82, Setd1B and B-676 over-expression had fewer viable cells than Setd1A, suggesting a functional difference between the two methyltransferases independent of interaction with Wdr82. Setd1B truncations B-2482 through B-4981 were not statistically different than Setd1A, suggesting a unique binding motif is located between nucleotides 676 and 2482 of Setd1B cDNA. This finding suggests structural differences in the central region of the two methyltransferases play an important role in functionality, consistent with the HCF-1 binding site found in the central region of Setd1A, but not Setd1B. HCF-1 is a transcriptional co-activator that interacts with Mll1 in addition to Setd1A, and is required for targeting these methyltransferases to IE62 target genes important for herpes simplex virus activation (Narayanan et al., 2007). HCF-1 plays a significant role in the G1 to S transition. For example, G1 to S transition is enabled with removal of a repressive H4K20 mono-methylation mark by PHF8. PHF8 is recruited to E2F1 regulated promoters based on interaction with di- and trimethylated H3K4 in conjunction with E2F1, HCF-1 and Setd1A (Liu et al., 2010). However, both Setd1A and Setd1B over-expression resulted in a higher percentage of cells in G1 phase and fewer in S phase compared to vector, yet not significantly different from each other as observed with PI

cell cycle analysis. This confirms both Setd1 proteins play a role in the cell cycle, but the mechanism for differences observed in cell numbers cannot be identified. It is likely that apoptosis plays a role, and can be evaluated with several different methods including evaluation of the fragmentation of DNA in which late stage apoptotic cells break into different length pieces (TUNEL assay). Alternatively, assays for alteration in membrane asymmetry with phosphatidylserine translocation from cytoplasmic to extracellular side of the cell membrane, or activation of apoptotic caspases for early apoptotic events can be pursued.

Previously, our lab showed that over-expression of the carboxyl-terminal fragment of either Setd1A or Setd1B resulted in lower levels of both Setd1 proteins (Lee et al., 2007). It is unknown what impact over-expression of full length or partial Setd1B fragments has on endogenous levels of either Setd1A or Setd1B. It is possible that over-expression of Setd1B truncations actually reduce Setd1A and/or Setd1B levels similar to the observation with the small carboxyl-terminal fragments, and this needs to be carefully characterized. Nonetheless, over-expression of either full length Setd1 protein resulted in decreased number of viable cells compared to vector, which suggests increased levels of Setd1 alters cell division dynamics. Since an increase in apoptosis and slow growth was observed with reduced Setd1A levels (Tate et al., 2009b), it is possible that the level of Setd1 proteins play a role in cell cycle regulation. Alternatively, the relative levels of Setd1A to Setd1B may impact cell cycle progression.

Although growth characteristics were not evaluated, reduction of Setd1B by siRNA did not reduce levels of Setd1A in NIH 3T3 cells (Wu et al., 2008). Unfortunately, levels of Setd1B could not be evaluated due to lack of antibody. It is

interesting to note that lowered Setd1A levels accompany reduced H3K4 trimethylation in siRNA Wdr82 depleted cells which is not replicated by siRNA reduction of Setd1A or Setd1B alone. The reduced level of H3K4 trimethylation is replicated when both Setd1 proteins are reduced together. Increased levels of H3K4 trimethylation are observed in Cfp1 null ES cells despite lowered Setd1A levels, suggesting Setd1B may be responsible. (Wu et al., 2008; Tate et al., 2010). A better understanding of the relationship between Setd1A and Setd1B levels might help further identify the functional role of these methyltransferases.

### 4.3 Cfp1 is a critical regulator of transcription

We have previously shown that ES cells lacking Cfp1 have increased levels of global H3K4 trimethylation with mislocalization to areas of heterochromatin by confocal analysis. Although H3K4 trimethylation is primarily associated with active transcription, microarray analysis demonstrated that altered H3K4 trimethylation results in increased and decreased transcription of genes, indicating a role in transcriptional repression as well. This is consistent with the finding that H3K4 trimethylation is required for silencing of telomeres, rDNA , HML and HMR loci in *Saccharomyces cerevisiae* (Fingerman et al., 2005). Consistently, a larger number of genes were up-regulated in Cfp1 null ES cells than down-regulated, suggesting the repressive role of Cfp1 is not minimal.

The large numbers of genes differentially expressed between ES cells with and without Cfp1 clearly implicate Cfp1 as a critical regulator of gene expression. The large difference between the two nulls is suggestive of primary and secondary effects from the

absence of Cfp1. Consistent mis-regulation and rescue most likely reveal Cfp1 targeted genes, and identification of these genes can be utilized for future study. However, the lack of complete recovery in Rescue cell line may suggest absence of Cfp1 results in a sustainably altered epigenetic program at some loci. Altered epigenetic programming preventing Cfp1 interaction may contribute to differences observed between the two Cfp1 nulls or alternatively, may be the random signaling due to lack of tight regulation of cellular processes in absence of Cfp1. Exploration of specific loci that did not rescue upon re-introduction of Cfp1 should be performed to assess epigenetic marks including DNA methylation status.

Cfp1 interacts with Setd1A, Setd1B, Mll1 and Mll2 and changes observed in transcription are likely due to altered functionality of these methyltransferases in addition to changes in DNA methylation (Dehé et al., 2006; Ansari et al., 2008; Wu et al., 2008, Lee et al., 2007). Evidence suggests the majority of global H3K4 trimethylation is attributed to Setd1A as opposed to Mll methyltransferases, however, the role of Setd1B has not been sufficiently evaluated (Wu et al., 2008; Wang et al., 2009). Setd1A levels decrease in the absence of Cfp1 and the methyltransferases(s) responsible for increased H3K4 trimethylation is unknown. The list of aberrant genes rescued with re-introduction of Cfp1 should provide a targeted focus for investigation of methyltransferase occupancy by chromatin immunoprecipitation at specific loci, eliminating the need for genome wide investigation. Identification of specific methyltransferases at targeted genes should help explain the role of multiple H3K4 specific methyltransferases interacting with Cfp1.

### 4.4 Cfp1 regulated transcription impacts multiple basic cellular processes

Genes in every chromosome examined, including the X chromosome exhibited altered levels of transcription in absence of Cfp1 and were associated with a broad range of biological processes. DAVID functional analysis associated up and down regulated genes with primary processes, including key developmental genes and those involved in apoptosis, consistent with phenoytpe defects observed in absence of Cfp1. Specifically, genes that require tight regulation during development such as Fgf4 and Suz12, were up-regulated in addition to fifteen homeobox containing genes, confirming Cfp1 plays a critical role in regulating differentiation and development. Notably, Xist which is important for X-inactivation was down-regulated and rescued suggesting H3K4 methylation plays a role in maintaining X-inactivation critical for development. Over 20 genes involved in programmed cell death were identified, including decreased BID and Bnip3L and increased Cidea (cell death inducing DNA fragmentation factor), consistent with the previous finding of increased apoptosis in Cfp1 depleted cells (Carlone & Skalnik, 2001; Tate et al., 2009b). Several genes implicated in DNA repair were mis-regulated, including the down regulation of XPA, consistent with the finding of increased susceptibility to DNA damage in Cfp1 null ES cells (Tate et al., 2009a). Not surprisingly, the largest group of mis-regulated genes was associated with nucleic acid interacting proteins indicating a critical link between Cfp1 and DNA and RNA machinery. Notably, many of the up-regulated genes facilitate post-translational modifications necessary for signaling events in molecular pathways and development and it is difficult to discern if these are a primary or secondary effect from loss of Cfp1. A correlation between mis-regulated genes and CGI promoters could strengthen the case for

direct correlation with Cfp1, as Cfp1 was found highly enriched at CGI sites (Thomson et al., 2010).

In conclusion, these studies show a critical link between Cfp1 and H3K4 trimethylation to regulation of basic cellular processes. The interaction of Cfp1 and Wdr82 with both common and distinct methyltransferases calls for a closer look at those targeted to specific genes, along with clarification of their roles in activation or suppression of transcription.

## REFERENCES


Akashi, K., X. He, J. Chen, H. Iwasaki, C. Niu, B. Steenhard, J. H. Zhang, J. & L. Li, (2003) Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**: 383-390.

Andreu-Vieyra, C. V., R. Chen, J. E. Agno, S. Glaser, K. Anastassiadis, A. F. Stewart & M. M. Matzuk, (2010) MLL2 Is Required in Oocytes for Bulk Histone 3 Lysine 4 Trimethylation and Transcriptional Silencing. *PLoS Biol* **8**: e1000453.

Ansari, K. I., B. P. Mishra & S. S. Mandal, (2008) Human CpG binding protein interacts with MLL1, MLL2 and hSet1 and regulates Hox gene expression. *Biochim. Biophys. Acta Gene Reg. Mech.* **1779**: 66-73.

Antequera, F. & A. Bird, (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA* **90**: 11995-11999.

Attema, J. L., P. Papathanasiou, E. C. Forsberg, J. Xu, S. T. Smale & I. L. Weissman, (2007) Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. *Proc. Natl. Acad. Sci. USA* **104**; 12371-12376

Ayton, P., S. F. Sneddon, D. B. Palmer, I. R. Rosewell, M. J. Owen, B. Young, R. Presley & V. Subramanian, (2001) Truncation of the MLL gene in exon 5 by gene targeting leads to early preimplantation lethallity of homozygous embryos. *Genesis* **30**: 201-212.

Bartolomei, M. S. & S. M. Tilghman, (1997) Genomic imprinting in mammals. *Annu. Rev. Genet.* **31**: 493-525.

Baur, A. S., P. Shaw, N. Burri, F. Delacretaz, F. T. Bosman & P. Chaubert, (1999) Frequent methylation silencing of p15(INK4b)(MTS2) and p16(INK4a)(MTS1) in B-cell and T-cell lymphomas. *Blood* **94**: 1773-1781.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber & E. S. Lander, (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.

Bestor, T. H., (2000) The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9**: 2395-2402.

Bird, A., (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**: 6-21.

Boa, S., C. Coert & H. G. Patterton, (2003) Saccharomyces cerevisiae Set1p is a methyltransferase specific for lysine 4 of histone H3 and is required for efficient gene expression. *Yeast* **20**: 827-835.

Briggs, S. D., M. Bryk, B. D. Strahl, W. L. Cheung, J. K. Davie, S. Y. R. Dent, F. Winston & C. D. Allis, (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev.* **15**: 3286-3295.

Butler, J. S., J.-H. Lee & D. G. Skalnik, (2008) CFP1 Interacts with DNMT1 independently of association with the Setd1 histone H3K4 methyltransferase complexes. *DNA Cell Biol.* **27**: 533-543.

Butler, J. S., L. R. Palam, C. M. Tate, J. R. Sanford, R. C. Wek & D. G. Skalnik, (2009) DNA methyltransferase protein synthesis is reduced in CXXC finger protein 1 deficient embryonic stem cells. *DNA Cell Biol.* **28**: 223-231.

Carlone, D. L., J.-H. Lee, S. R. L. Young, E. Dobrota, J. S. Butler, J. Ruiz & D. G. Skalnik, (2005) Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Mol. Cell. Biol.* **25**: 4881-4891.

Carlone, D. L. & D. G. Skalnik, (2001) CpG binding protein is crucial for early embryonic development. *Mol. Cell. Biol.* **21**: 7601-7606.

Cheng, H., X. He & C. Moore, (2004) The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. *Mol. Cell. Biol.* **24**: 2932-2943.

Corda, Y., V. Schramke, M. Pia Longhese, T. Smokvina, V. Paciotti, V. Brevet, E. Gilson & V. Geli, (1999) Interaction between Sep1p and checkpoint protein Mec3p in DNA repair and telomere functions. *Nature Genet.* **21**: 204-208.

Dehé, P.-M., B. Dichtl, D. Schaft, A. Roguev, M. Pamblanco, R. g. Lebrun, A. Rodríguez-Gil, M. Mkandawire, K. Landsberg, A. Shevchenko, A. Shevchenko, L. E. Rosaleny, V. Tordera, S. N. Chávez, A. F. Stewart & V. Géli, (2006) Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J. Biol. Chem.* **281**: 35404-35412.

Dehé, P.-M., M. Pamblanco, P. Luciano, R. Lebrun, D. Moinier, R. Sendra, A. Verreault, V. Tordera & V. Géli, (2005) Histone H3 lysine 4 mono-methylation does not require ubiquitination of histone H2B. *J. Mol. Biol.* **353**: 477-484.

Dichtl, B., R. Aasland & W. Keller, (2004) Function for S. cerevisiae Swd2p in 3' end formation of specific mRNAs and snRNAs and global histone 3 lysine 4 methylation. *RNA* **10**: 965-977.

Felsenfeld, G. & M. Groudine, (2003) Controlling the double helix. *Nature* **421**: 448-453.

Fingerman, I. M., C.-L. Wu, B. D. Wilson & S. D. Briggs, (2005) Global Loss of Set1-mediated H3 Lys4 Trimethylation is associated with silencing defects in Saccharomyces cerevisiae. *J. Biol. Chem.* **280**: 28761-28765.

Glaser, S., J. Schaft, S. Lubitz, K. Vintersten, F. van der Hoeven, K. R. Tufteland, R. Aasland, K. Anastassiadis, S.-L. Ang & A. F. Stewart, (2006) Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**: 1423-1432.

Grewal, S. I. & D. Moazed, (2003) Heterochromatin and epigenetic control of gene expression. *Science* **301**: 798-802.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch & R. A. Young, (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**: 77-88.

Guo, S.-X., T. Taki, H. Ohnishi, H.-Y. Piao, K. Tabuchi, F. Bessho, R. Hanada, M. Yanagisawa & Y. Hayashi, (2000) Hypermethylation of p16 and p15 genes and RB protein expression in acute leukemia. *Leuk. Res.* **24**: 39-46.

Hennessy, B. T., G. Garcia-Manero, H. M. Kantarjian & F. J. Giles, (2003) DNa methylation in haematological malignancies: the role of decitabine. *Expert Opin. Investig. Drugs* **12**: 1985-1993.

Hwang, W. W., S. Venkatasubrahmanyam, A. G. Ianculescu, A. Tong, C. Boone & H. D. Madhani, (2003) A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol. Cell* **11**: 261-266.

Issa, J. P., B. A. Zehnbauer, S. H. Kaufmann, M. A. Biel & S. B. Baylin, (1997) HIC1 hypermethylation is a late event in hematopoietic neoplasms. *Cancer Res.* **57**: 1678-1681.

Jaenisch, R. & A. Bird, (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet. Suppl.* **33**: 245-254.

Jenuwein, T. L., G.; Dorn, R.; Reuter, G., (1998) SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci.* **54**: 80-93.

Kanehisa, M. & S. Goto, (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**: 27-30.

Kleinjan, D. A., A. Seawright, A. J. Childs & V. van Heyningen, (2004) Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. *Dev. Biol.* **265**: 462-477.

Krogan, N. J., J. Dover, S. Khorrami, J. F. Greenblatt, J. Schneider, M. Johnston & A. Shilatifard, (2002) COMPASS, a histone H3 (lysine 4) methyltransferase required for telomeric silencing of gene expression. *J. Biol. Chem.* **277**: 10753-10755.

Krogan, N. J., J. Dover, A. Wood, J. Schneider, J. Heidt, M. A. Boateng, K. Dean, O. W. Ryan, A. Golshani, M. Johnston, J. F. Greenblatt & A. Shilatifard, (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell* **11**: 721-729.

Lander, E. S., (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Larsen, F., G. Gundersen, R. Lopez & H. Prydz, (1992) CpG islands as gene markers in the human genome. *Genomics* **13**: 1095-1107.

Lee, J.-H. & D. G. Skalnik, (2005) CpG binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* **280**: 41725-41731.

Lee, J.-H. & D. G. Skalnik, (2008) Wdr82 is a CTD-binding protein that recruits the Setd1A histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol. Cell. Biol.* **28**: 609-618.

Lee, J.-H., J. You, E. Dobrota & D. G. Skalnik, (2010) Identification and characterization of a novel human PP1 phosphatase complex. *J. Biol. Chem.* **285**: 24466-24476.

Lee, J. H., C. M. Tate, J. S. You & D. G. Skalnik, (2007) Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J. Biol. Chem.* **282**: 13419-13428.

Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch & E. Li, (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**: 3195-3205.

Liu, W., T. Bogdan, V. T. Oksana, T. Y. Zhou, R. Gassmann, W.T. Liu, K. A. Ohgi, C. Benner, I. Garcia-Bassests, A. K. Aggarwal, A. Desai, P. C. Dorrestein, C.K. Glass & M. G. Rosenfeld, (2010) PHF8 mediates hitone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**: 508-512

Macleod, D., R. R. Ali & A. Bird, (1998) An alternative promoter in the mouse major histocompatibility complex class II I-Abeta gene: implications for the origin of CpG islands. *Mol. Cell. Biol.* **18**: 4433-4443.

McClintick, J. & H. Edenberg, (2006) Effects of filtering by Present call on analysis of microarray experiments. *BMC Bioinformatics* **7**: 49.

Miller, T., N. J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J. F. Greenblatt & A. Shilatifard, (2001) COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. USA* **98**: 12902-12907.

Morillon, A., N. Karabetsou, A. Nair & J. Mellor, (2005) Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Mol. Cell* **18**: 723-734.

Mueller, C. L. & J. A. Jaehning, (2002) Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol. Cell. Biol.* **22**: 1971-1980.

Nagy, P. L., J. Griesenbeck, R. D. Kornberg & M. L. Cleary, (2002) A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. *Proc. Natl. Acad. Sci. USA* **99**: 90-94.

Narayanan, A., W. t. Ruyechan & T. M. Kristie, (2007) The coactivator host cell factor-1 mediates Set1 and MLL1 H3K4 trimethylation at herpesvirus immediate early promoters for initiation of infection. *Proc. Natl. Acad. Sci. USA* **104**: 10835-10840.

Ng, H.-H., F. Robert, R. A. Young & K. Struhl, (2003a) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11**: 709-719.

Ng, H. H., S. Dole & K. Struhl, (2003b) The Rtf1 Component of the Paf1 Transcriptional Elongation Complex Is Required for Ubiquitination of Histone H2B. *J. Biol. Chem.* **278**: 33625-33628.

Nislow, C., E. Ray & L. Pillus, (1997) SET1, a yeast member of the Trihorax family, functions in transcriptional silencing and diverse cellular processes. *Mol. Biol. Cell* **8**: 2421-2436.

Okano, M., D. W. Bell, D. A. Haber & E. Li, (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.

Okano, M. & E. Li, (2002) Genetic analyses of DNA methyltransferase genes in mouse model system. *J. Nutr.* **132**: 2462S-2465.

Pokholok, D. K., C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford & R. A. Young, (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**: 517-527.

Roguev, A., D. Schaft, A. Shevchenko, R. Aasland, A. Shevchenko & A. F. Stewart, (2003) High conservation of the Set1/Rad6 axis of histone 3 lysine 4 methylation in budding and fission yeasts. *J. Biol. Chem.* **278**: 8487-8493.

Roguev, A., D. Schaft, A. Shevchenko, W. Pim Pijnappel, M. Wilm, R. Aasland & A. F. Stewart, (2001) The Saccharomyces cerevisiae Set1 complex includes an ASH2 homologue and methylates histone lysine 4. *EMBO J.* **20**: 7137-7148.

Roguev, A., A. Shevchenko, D. Schaft, H. Thomas, A. F. Stewart & A. Shevchenko, (2004) A comparative analysis of an orthologous proteomic environment in the Yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Mol . Cell. Proteomics* **3**: 125-132.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor & T. Kouzarides, (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* **419**: 407-411.

Schneider, J., A. Wood, J. S. Lee, R. Schuster, J. Dueker, C. Maguire, S. K. Swanson, L. Florens, M. P. Washburn & A. Shilatifard, (2005) Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol. Cell* **19**: 849-856.

Shahbazian, M. D., K. Zhang & M. Grunstein, (2005) Histone H2B Ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. *Mol. Cell* **19**: 271-277.

Skarnes, W. C., J. E. Moss, S. M. Hurtley & R. S. Beddington, (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl. Acad. Sci. USA* **92**: 6592-6596.

Storey, J. D. & R. Tibshirani, (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**: 9440-9445.

Takeda, S., D. Y. Chen, T. D. Westergard, J. K. Fisher, J. A. Rubens, S. Sasagawa, J. T. Kan, S. J. Korsmeyer, E. H. Y. Cheng & J. J. D. Hsieh, (2006) Proteolysis of MLL family proteins is essential for Taspase1-orchestrated cell cycle progression. *Genes Dev.* **20**: 2397-2409.

Tate, C. M., M. L. Fishel, J. L. Holleran, M. J. Egorin & D. G. Skalnik, (2009a) Embryonic stem cells lacking the epigenetic regulator Cfp1 are hypersensitive to DNA-damaging agents and exhibit decreased Ape1/Ref-1 protein expression and endonuclease activity. *DNA Repair* **8**: 1411-1423.

Tate, C. M., J.-H. Lee & D. G. Skalnik, (2009b) CXXC Finger Protein 1 Contains redundant functional domains that support embryonic stem cell cytosine methylation, histone methylation, and differentiation. *Mol. Cell Biol*. **29**: 3817-3831.

Tate, C. M., J. H. Lee & D. G. Skalnik, (2010) CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. *FEBS Journal* **277**: 210-223.

Tenney, K. & A. Shilatifard, (2005) A COMPASS in the voyage of defining the role of trithorax/MLL-containing complexes: Linking leukemogensis to covalent modifications of chromatin. *J. Cell. Biochem.* **95**: 429-436.

Thomson, J. P., P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. W. Kerr, A. Deaton, R. Andrews, K. D. James, D. J. Turner, R. Illingworth & A. Bird, (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**: 1082-1086.

Tyagi, S., A. L. Chabes, J. Wysocka & W. Herr, (2007) E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol. Cell* **27**: 107-119.

Voo, K. S., D. L. Carlone, B. M. Jacobsen, A. Flodin & D. G. Skalnik, (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* **20**: 2108-2121.

Wang, P., C. Lin, E. R. Smith, H. Guo, B. W. Sanderson, M. Wu, M. Gogol, T. Alexander, C. Seidel, L. M. Wiedemann, K. Ge, R. Krumlauf & A. Shilatifard, (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional Initiation by RNA polymerase II. *Mol. Cell. Biol.* **29**: 6074-6085.

Wu, M., P. F. Wang, J. S. Lee, S. Martin-Brown, L. Florens, M. Washburn & A. Shilatifard, (2008) Molecular regulation of H3K4 Trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol. Cell Biol.* **28**: 7337-7344.

Xu, G.-L., T. H. Bestor, D. Bourc'his, C.-L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russso & E. Viegas-Pequignot, (1999) Chromosome instability and immunodefiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**: 187-191.

Yoder, J. A., C. P. Walth & T. H. Bestor, (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**: 335-340.

Yokochi, T. & K. D. Robertson, (2002) Preferential methylation of unmethylated DNA by mammaliande Novo DNA methyltransferase Dnmt3a. *J. Biol. Chem.* **277**: 11735-11745.

Zhu, J., F. He, S. Hu & J. Yu, (2008) On the nature of human housekeeping genes. *Trends Genet.* **24**: 481-484.

# CURRICULUM VITAE

## PATRICIA A. PICK-FRANKE

### EDUCATION

M.S. Biochemistry and Molecular Biology 2010
Indiana University,
Indianapolis, Indiana
Dissertation: Setd1 Histone H3K4 Methyltransferase Complex Components in Epigenetic Regulation. Advisor, Dr. D. Skalnik

B.S. Animal Science 1985 *Cum Laude*
Washington State University
Pullman, Washington

### EXPERIENCE

1987-2003 Sales and Marketing Executive for medical diagnostic companies
1986-1987 Research Scientist- Hycor Biomedical Garden Grove, California
Commercial development of ELISA for infectious diseases
1984-1986 Research Scientist- Washington State University Pullman, Washington
Monoclonal antibody generation, protein purification and immunoassay development

### SKILLS

Technical skills: Assay development, PCR, western blot, flourescent microscopy, confocal microscopy, mouse breeding, dissection and handling, stem cell culture, tissue culture techniques, microarray, protein purification, genetic modification and cloning
Computer skills: Microsoft office suite, Adobe suite

### PUBLICATIONS

Bhat-Nakshatri, P., Appaiah, H., Ballas, C., **Pick-Franke, P**., Goulet, R., Badve, S., Srour, E., and Nakshatri, H., (2010) SLUG/SNAI2 and Tumor Necrosis Factor Generate Breast Cells With CD44+/CD24- Phenotype. *BMC Cancer* **10**: 411.